Performance Measures for the Federal Investment in Environmental and Natural Resources Research and Development

Robert J. Lempert, Beth E. Lachman, Susan Fuhs, and N. Darius Sankey

DRU-1179-CTI
August 1995
Prepared for the Critical Technologies Institute

The RAND unrestricted draft series is intended to transmit preliminary results of RAND research. Unrestricted drafts have not been formally reviewed or edited. The views and conclusions expressed are tentative. A draft should not be cited or quoted without permission of the author, unless the preface grants such permission.

RAND is a nonprofit institution that seeks to improve public policy through research and analysis. RAND's publications and drafts do not necessarily reflect the opinions or policies of its research sponsors.
TABLE OF CONTENTS

Preface .................................................................................................................. i
Summary ............................................................................................................... iii

1. Introduction .................................................................................................... 1

2. Linking Programs to Goals ........................................................................... 7

3. Performance Measures for CENR ............................................................... 15
   3.1 Measures for Individual Programs .......................................................... 17
   3.2 Measures for Program Balance ............................................................... 27

4. Using the Template ....................................................................................... 33

Bibliography ....................................................................................................... 39

Tables .................................................................................................................. 41

Figures ................................................................................................................ 44

Appendices
   A. August 1994 Version of the Template ................................................... 47
   B. Industrial Research Institute (IRI) Metrics .............................................. 59
PREFACE

This report describes a template -- a 'how to' guide and reference manual -- designed to help the Committee on Environment and Natural Resources (CENR) of the National Science and Technology Council (NSTC) develop performance measures for their research and development (R&D) programs. Performance measures have become increasingly important as the political leadership in the Congress and Executive Branch have directed government agencies of all types to quantitatively evaluate the effectiveness of their expenditures of public funds.

This performance measures template was developed by the Critical Technologies Institute (CTI) for the NSTC and the Office of Science and Technology Policy. The work is part of a larger CTI inquiry into evaluating the performance of government research programs, including an examination of metrics for fundamental science conducted for the NSTC's Science Subcommittee on Research (NSTC/CFS/SR) and a review of industry methods of evaluating research, conducted for the NSTC's Committee on Civilian Industrial Technologies (NSTC/CIT).

CTI developed the performance measures template described in this report in close consultation with CENR during the committee's preparations for its FY96 budget submission during the summer and fall of 1994. We began our work in June, and in July presented drafts to OSTP and two CENR subcommittees, Global Change Research (SGCR) and Biodiversity and Ecosystem Dynamics (SBED). We distributed an initial template to the full CENR in August and helped the SGCR and SBED subcommittees use it in developing their strategic and implementation plans. Our work with CENR ended in October 1994, when the subcommittees completed their draft plans.

This report presents a revised version of our August 1994 template. We intend that in the future CENR, as well as managers of other government R&D programs, can use this template to design more effective performance measures for their programs.
Created in 1992 by an act of Congress, CTI is a federally funded research and development center house within RAND. CTI's missions is to:

- Provide analytical support to the Executive Office of the President of the United States, Office of Science and Technology Policy (OSTP);

- Help decisionmakers understand the likely consequences of their decisions and choose among alternative policies; and

- Improve understanding in both the public and private sectors of ways in which technological efforts can better serve national objectives.

For more information on CTI or this document, please contact:

Stephen M. Drezner
Director, Critical Technologies Institute
RAND
2100 M Street, N.W.
Washington, D.C. 20037
202-296-5000
fax: 202-452-8377

or see the CTI Homepage at:

www.rand.org:80/centers/cti/
SUMMARY

Over the last several years, there has been increasing pressure to make government programs of all types more effective and more accountable to the public and their representatives. In particular, the Executive Branch and Congress have directed managers of federal research programs to quantitatively assess, measure, and otherwise describe the outcomes of their programs for the purposes of managing them more effectively and of allocating funds among different research endeavors.

Despite these calls for accountability, the difficulties inherent in measuring the performance of research and development programs are also widely recognized. Research has fundamentally unpredictable outcomes, and the outcomes of particular research efforts often have unpredictable uses. It is common to find seemingly esoteric discoveries which, sometimes years later, provide the critical information needed to complete some development with profound benefits for the US economy, national security, human health, or the environment. Many commentators have voiced fears that because of the great difficulties in developing effective performance measures for research, government agencies will adopt inappropriate measures which may stifle the creativity of researchers, waste the time and financial resources of the research community, direct researchers towards less useful but more easily measured activities, and inappropriately shift resources away from highly useful activities.

These potential problems are particularly acute for federal investments in environmental and natural resources research and development. The National Science and Technology Council's Committee on Environment and Natural Resources (CENR), has authority over many types of R&D activities, ranging across fundamental science, technology development, data collection, and integrated assessments to help decisionmakers in making appropriate environmental and natural resources policy decisions. CENR must develop performance measures
which help ensure an efficient and effective effort across this broad and diverse range of activities.

This report describes a template designed to help CENR develop performance measures for the research and development activities under their purview. The template was designed as a 'how-to' guide and reference manual to help the CENR prepare their environment and natural resources R&D strategic and implementation plans for their fiscal year 1996 budget requests. The template employs three central concepts to address the common problems with evaluating research programs:

- Performance measures are most effective when they refer to a specific set of goals for the research. The template presents a four level hierarchy of goals designed to help CENR carefully define their objectives and to trace the links between the goals of individual R&D programs and broad national goals.

- Different types of research and development demand different types of performance measures. For instance, measures appropriate for technology development programs are not usually appropriate for programs in basic science. The template helps CENR systematically differentiate its various types of R&D and find the most effective set of metrics for each.

- The unpredictability of research and the diverse range of environmental R&D needs strongly suggest that CENR view its R&D programs as a portfolio of activities. A portfolio spreads risks over a number of investments. While any single investment may have a high chance of failing to meet its goals, the portfolio as a whole can be highly likely to succeed. The template provides performance measures of program balance which CENR can use to complement the measures of individual programs generally envisioned by the political leadership. For instance, balance measures can help CENR find an appropriate balance of spending between long-term,
potentially high risk but high payoff basic research and short-term efforts aimed at providing answers to specific policy decisions.

In order to implement these concepts, the template provides a framework for developing a hierarchy of goals, choosing performance measures for different types of individual R&D programs, and choosing performance measures of program balance across different programs. This report also provides a long list of candidate performance measures which CENR can consider for possible use. These candidate measures are divided into categories corresponding to each part of the template. The measures are drawn from the experience of government managers of fundamental science programs and from the experience of industry managers of applied research. This report gives enough description of each measure so that the members of CENR can make an initial screening of those they find most interesting. We then refer the reader to specific literature for a fuller discussion of any measure they are seriously considering using in their strategic and implementation plans.

The template advises the CENR to follow a five step process in developing performance measures for its R&D programs:

1) Develop a specific hierarchy of goals relating broad national goals, to environmental policy goals, to R&D goals, to the goals of individual CENR research and development programs.

2) Divide the R&D programs into different categories of activities, each appropriate to a different set of performance measures. We suggest the categories of basic science, technology development, data collection, and assessment. However, some CENR subcommittees may find other categories more appropriate.

3) Develop performance measures for individual R&D programs to help evaluate how well programs are managed, and in
some cases, how well they contribute to higher level goals. We provide a menu of candidate measures.

4) Develop measures of program balance to help evaluate the choice and mix of different types of CENR R&D activities and to help evaluate the contribution of these programs to higher level national goals. We provide a menu of candidate measures.

5) Use the performance measures to monitor the progress of CENR R&D activities, and to modify these activities as necessary as more is learned about the efficiency of various programs, the promise of the research they support, and their usefulness to CENR's and the nation's goals.

Our work with the CENR during its preparations for the 1996 budget suggests several lessons which the committee should keep in mind while using the template in future years. These lessons emphasize that developing performance measures will be a long-term process of experimentation and learning.

First, it is important that CENR make performance measures part of its planning processes from the very beginning, expressing the goals of its programs in such a way as to facilitate the use of performance measures and providing CENR managers and planners sufficient time to shape their programs to succeed in the criteria being measured.¹

Second, many of the potentially most promising performance measures will require CENR to gather significant amounts of data -- including detailed budget information on CENR and other comparable R&D programs, and information on the impacts of CENR R&D -- to establish benchmarks for the balance among CENR programs and the performance of individual programs. Thus it may take several years

¹ Otherwise, CENR will tend to choose measures which flatter the current set of programs, rather than those which would prove most beneficial to their research endeavor in the long-term.
before CENR develops a fully effective set of performance measures, since much of this data collection will require the establishment of multi-year trends.

Finally, CENR should not try to use too many performance measures for evaluating its programs. It is not yet clear to us how many is too many, and CENR should probably experiment with a number of different measures for several years before deciding which it finds most useful.

Performance measures will remain a significant challenge for CENR and other federal R&D programs. There are significant dangers to incomplete, excessive, or poorly conceived performance measures. However, the demands for accountability in federal R&D are likely to remain strong. We believe that this template can help guide CENR through the process of developing effective performance measures for their programs, provide a menu of options for choosing specific measures, and develop an evaluation process consistent with a creative and valuable R&D enterprise and the demands of the accountability to the public and its representatives.
1. INTRODUCTION

The Federal government spends roughly $5.3 billion [NSTC, 1995] annually on environmental and natural resources research and development. Some of this effort is basic science, meant to advance our fundamental understanding of the interactions between society and nature; some is technology R&D, intended to improve society's options for improving environmental quality; and some is designed to help government, and increasingly, private sector, decision-makers craft more effective policies to protect the environment and human health. Like most federal programs, environmental and natural resources R&D is under increased pressure to become more accountable to the taxpayers and their representatives. This interest is encouraged by a number of trends, including attempts to balance the federal budget, growing demands in the number and expense of potentially important R&D programs, and a general desire to make government more accountable to the people.

Over the last several years such pressures have coalesced into specific legislation and executive initiatives which require that all federal programs, including federally-supported R&D, provide clear linkages to societal goals and provide performance measures which can show progress towards these goals. The legislation and executive initiatives include the National Performance Review (NPR) [Gore, 1993] and the Government Performance and Results Act (GPRA) of 1993 (PL 103-62). These directives emphasize two themes. First, all government programs should be clearly related to specific national goals so that administrators and taxpayers know what ends each program aims to achieve. Second, the performance of each program should wherever possible be measured quantitatively on the basis of program outcomes rather than inputs.

It is widely recognized that the managers of federal R&D will have a difficult time developing quantitative measures for their programs. Research has fundamentally unpredictable outcomes, and the outcomes of particular research efforts often have unpredictable uses. It is common to find seemingly esoteric discoveries which, sometimes years later, provide the critical information needed to complete some development
with profound benefits for the US economy, national security, human health, or the environment. Not surprisingly, numerous critiques of performance measures for scientific research show that none will entirely satisfy the growing demands for quantitative accountability [Wagner and Flanagan, 1995].

In addition, many commentators have also voiced fears that because of the great difficulties in developing effective performance measures for R&D, government agencies will adopt inappropriate measures which may pose dangers to the post-war research enterprise that in many respects has performed extraordinarily well over the last 40 years [Ziman, 1994]. Research and development is an activity which thrives on creativity and individual initiative. Performance measures may waste the time and energy of the R&D community; may direct researchers towards less useful activities (e.g. publishing many substandard papers to increase publications counts); and inappropriately shift resources towards those activities more easily measured (e.g. information dissemination activities) and away from those more difficult to measure (e.g. fundamental research).

This report describes a set of concepts and procedures designed to help the Committee on Environmental and Natural Resources (CENR) of the National Science and Technology Council (NSTC) develop effective and beneficial performance measures for the R&D under their purview. The NSTC is a Cabinet-level, interagency council whose objectives include establishing clear national goals for federal science and technology investments and coordinating the science and technology policymaking and implementation process across federal agencies. A primarily responsibility of the NSTC is to develop strategic plans for federal research which articulate the goals, objectives, milestones, metrics for the Federal research system. [NSTC, 1995]

The CENR is one of nine NSTC committees, each covering a broad area of federal research and development, as shown in Figure 1. The CENR coordinates interagency programs among 12 agencies and focuses environmental and natural resources R&D on those questions that most
directly impact our health and our economy. [NSTC, 1995] Perhaps more than any other NSTC committee, CENR faces great hurdles in designed performance measures. CENR has authority over many types of research and development activities, ranging across fundamental science, technology development, data collection, and research which helps address specific environmental policy decisions. In addition, CENR has a diverse set of goals for its R&D. Like the fundamental science programs under the purview of the NSTC Committee on Fundamental Science (CFS), much of CENR's activities aim to increase our understanding of the world around us. But like the applied research and development under the purview of the Committee on Industrial Technologies (CIT), much CENR research and development is directed to specific national goals. However, CENR's goals of enhancing environmental quality and improving the information available for environmental policy-making are more difficult to quantify and to compare to one another than CIT's goals of enhancing U.S. productivity and economic performance. CENR must develop performance measures which help ensure an efficient and effective effort across this broad and diverse range of activities.

The information in this report is presented in the form of a performance measures template. The template was designed as a 'how-to' guide and reference manual to assist the CENR prepare performance measures for its strategic and implementation plans for R&D which accompanied the Administration's fiscal year 1996 budget requests. In developing the template, we worked closely with the OSTP leadership of CENR and two of the CENR sub-committees: the Subcommittee on Biodiversity and Ecosystem Dynamics (SBED) and the Subcommittee on Global Change Research (SGCR) (see Figure 1). We chose to work with SBED and SGCR because they offered us the richest experience to incorporate into our template. Not only did the two subcommittees cover a broad range of research and development activities, they were also the subcommittees furthest advanced in developing their strategic and implementation plans when we began our work in June 1994. Our work with CENR and the two subcommittees ended in October 1994, when SBED and SGCR completed their draft plans.
The template described in this report draws primarily form two sources: our experience working closely with the CENR and the SBED and SGCR subcommittees, and companion studies conducted by CTI on performance measures for research. The first of these companion studies examined metrics for fundamental science for the NSTC Committee on Fundamental Science and the second examined industry metrics for applied research and development for the NSTC Committee on Civilian Industrial Technologies. An initial version of the template was distributed to the CENR in August 1994 (see Appendix A). We intend that the expanded and revised version of the template in this report will help explain to the larger community how CENR used performance measures in its current strategic and implementation plans, assist CENR in developing performance measures for future plans, and help lay the groundwork for the continued development and refinement of CENR performance measures in the years ahead.

The performance measures template employs three central concepts to address the common problems with evaluating R&D programs:

- Performance measures are most effective when they refer to a specific set of goals for the R&D activities. The template presents a four level hierarchy of goals designed to help CENR carefully define their objectives and to trace the links between the goals of individual R&D programs and broad national goals.

- Different types of research and development demand different types of performance measures. Measures appropriate for technology development programs are not appropriate for programs in basic science. The template helps CENR systematically differentiate its various types of R&D and find the most effective set of metrics for each.

- The unpredictability of research and the diversity of environmental and natural resources R&D needs suggest that CENR view its R&D programs as a portfolio of activities. A portfolio spreads risks over a number of investments. While any
single investment may have a high chance of failing to meet its
goals, the portfolio as a whole can be highly likely to succeed. The
template provides performance measures of program balance
which CENR can use to complement the measures of individual
programs generally envisioned by the political leadership. For
instance, balance measures can help CENR find an appropriate
balance of spending between long-term, potentially high risk but
high payoff basic research and short-term efforts aimed at
providing answers to specific policy decisions.

In order to implement these concepts, the template provides a
framework for developing a hierarchy of goals, choosing performance
measures for different types of individual R&D programs, and choosing
performance measures of program balance across different programs.
This report also provides a long list of candidate performance measures
which CENR can consider for possible use. These candidate measures
are divided into categories corresponding to each part of the template.
The measures are drawn from the experience of government managers of
fundamental science programs and from the experience of industry
managers of applied research and development. This report gives enough
description of each measure so that the members of CENR can make an
initial screening of those they find most interesting. We refer the reader
to specific literature for a fuller discussion of any measure they are
seriously considering using in their strategic and implementation plans.

The template advises the CENR to follow a five step process in
developing performance measures for its research and development
programs:

1) Develop a specific hierarchy of goals relating broad national
goals, to environmental policy goals, to research and development
goals, to the goals of individual CENR R&D programs.

2) Divide the research and development programs into different
categories of activities, each appropriate to a different set of
performance measures. We suggest the categories of basic
science research, technology development, data collection, and
assessments. However, some CENR subcommittees may find
other categories more appropriate.
3) Develop performance measures for individual research and development programs to help evaluate how well programs are managed, and in some cases, how well they contribute to higher level goals. We provide a menu of candidate measures in Table 2.

4) Develop measures of program balance to help evaluate the choice and mix of different types of CENR research and development activities and to help evaluate the contribution of these programs to higher level goals. In Table 4, we provide a menu of candidate measures.

5) Use the performance measures to monitor the progress of CENR research and development activities, and to modify these activities as necessary as more is learned about the efficiency of various programs, the promise of the R&D they support, and their usefulness to CENR's and the nation's goals.

The structure of this document reflects the steps we recommend CENR subcommittees should follow in developing performance measures for their R&D activities. Section 2 of this report describes a four-level hierarchy of goals which links the goals of individual CENR R&D programs to broad research and development goals. Section 3.1 provides a menu of performance measures for individual programs which CENR subcommittees can choose among to find those which best measure the performance of their programs. Section 3.2 provides a menu of performance measures of program balance which should be useful to CENR. The concluding Section 4 discusses how CENR used the concepts in this template in their current strategic plans and suggests some thoughts about how the template can be used in the future. In particular, we emphasize that developing effective performance measures for CENR should be a long-term process of experimentation and learning.
2. LINKING PROGRAMS TO GOALS

Performance measures are intended to link government programs to specific national goals. Thus, the first step in developing performance measures for CENR programs is clearly articulating the goals these programs intend to serve. The template presented here outlines a four-level hierarchy of goals which can help CENR and its subcommittees provide a firm structure for choosing an appropriate set of performance measures for their research and development activities. A hierarchy of goals is important because the different audiences who will use the performance measures, ranging from Congress to individual researchers, often focus on goals at different levels of specificity. In addition, a hierarchy allows a convenient separation among different types of CENR R&D. This is important because different types of R&D demand different types of performance measures.

CENR was created to improve the federal government's ability to conduct coordinated, cost-effective, interdisciplinary research to address key environmental issues [NSTC, 1995]. Many environmental issues pose multidimensional problems which can only be understood by bringing together researchers with many different areas of expertise -- natural scientists, social scientists, economists, and engineers. The traditional single agency and single discipline approach has often been insufficient to address society's environmental problems and has often left research decoupled from policy-makers. With representatives from all the federal agencies which fund environmental research and development, CENR aims to provide a coordinated, multi-agency, interdisciplinary approach to the planning and funding of these activities. By identifying redundancy in federal programs, filling critical research gaps, and anticipating the environmental problems of the future, CENR hopes to increase the productivity of the federal investment in environmental R&D and increase the contribution it makes to achieving the nations' environmental goals.

One of CENR's primary functions is to develop strategic and implementation plans for its research and development programs. These plans are meant to address several key audiences. CENR member
agencies such as the Department of Energy and NASA should use these CENR documents to better plan and manage their own R&D programs, so that the programs better reflect national goals and are more productive and efficient. The Office of Management and Budget will use the CENR plans to help allocate resources among programs based on the importance of the goals addressed and the effectiveness with which they are pursued. Similarly, Congress should use the CENR plans as a basis for allocating funds. Finally, stakeholder communities and the broad public can use the CENR plans to better understand the purposes for which the federal government is spending the public’s research investment, and to judge for themselves the utility of these investments.

The CENR uses its strategic and implementation plans to describe the relationship between federal investments in environmental and natural resource R&D and specific national goals. This key purpose flows directly from the requirements of the GPRA and the NPR. The public, Congress, program managers, and the agencies themselves require an explicit exposition of the relationship between goals and programs in order to share a clear idea of what the research and development programs intend to achieve. Because of the vast difference in the specificity of the goals which are useful to these very different audiences, it is useful to construct a hierarchy of goals, which relates the goals a R&D program manager might use to guide his or her day-to-day activities (for example, develop integrated assessments of climate change) to the broad national goals of most interest to the public and senior policymakers (for example, achieve sustainable development). Such hierarchies of goals are common in the strategic plans of many private sector firms and the military services. These hierarchies also help provide a framework for applying the performance measures for program balance (Section 3.2) which assess how well programs or groups of programs address goals at different levels.

During the budget process for fiscal year 1996, CENR gave its subcommittees guidance on writing their strategic and implementation plans for the research under their purview. CENR instructed its subcommittees to relate their R&D programs to specific national goals,
and to craft implementation plans which identified program goals and plans, programmatic milestones and performance measures, as well as the different agencies' roles in carrying out these plans.

In creating its template, CTI identified a four-level hierarchy of goals contained within the CENR guidance, shown in Figure 2. Broad national goals (level 1) are supported by environmental policy goals (level 2). These environmental policy goals are supported by goals for research and development (level 3), which in turn are supported by the goals and plans of individual research and development programs (level 4). We chose these four levels of goals to fit comfortably into the CENR guidance and minimize the subcommittees' work in using the template. Table 1 shows that the hierarchy in our template merely focuses on a four-level hierarchy implicit in the CENR guidance. We did, however, elaborate on the guidance in three ways. We have explicitly included broad national goals as the first level of the hierarchy although these goals will likely appear only implicitly in the subcommittees' strategic plans. Second, we also focused our attention on environmental policy goals within Level 2. Third, we refer to research and development, as opposed to just scientific, goals in Level 3, because the template will differentiate among different types of R&D and the appropriate performance measures for each (Section 3.1).

National goals are the highest level in our template's hierarchy. They are defined by the Carnegie Commission as "broad goals pursued mainly by one nation that derive from a domestic consensus on 'what is good for the country.'" [Carnegie Commission, 1992]. Generally senior level decision-makers articulate national goals they believe reflect public concerns. For instance, in developing its strategic and implementation plans, CENR was guided by the Administration's research Policy Principles, as set forth by Science Advisor John H. Gibbons and OMB director Leon Panetta in a May 6, 1994 memo to the heads of executive branch departments and agencies. Gibbons and Panetta identified six high level national goals to guide R&D investments in the FY 1996 budget:

- A Healthy, Educated Citizenry
• Job Creation and Economic Growth
• World Leadership in Science, Mathematics, and Engineering
• Improved Environmental Quality
• Harnessing Information Technology
• Enhanced National Security

The memo defined the goal of improved environmental quality as “protecting the environmental for present and future generation while enabling development to proceed in a sustainable manner. Environmental quality entails preserving ecological integrity and diversity, protecting human health and safety, and enhancing the quality and quantity of food, fiber, energy, and water supplies.” The memo states that the nation is committed to improving environmental quality at the local, regional, national, and global levels.

The nation pursues a number of environmental policies in order to achieve its broad goal of improved environmental quality. These environmental policy goals are the second level in our hierarchy. Environmental policy goals answer the question posed in the CENR guidance: “what actions do the federal government and other actors need to take to move us towards our national goals?” For example, goals can be embodied in international agreements such as the Montreal Protocol on Substances that Deplete the Ozone Layer, which calls for the elimination of chlorofluorcarbons (CFCs) and other ozone-depleting chemicals. Federal environmental goals are usually decided by Congress and the executive branch through the processes of legislation, regulatory rule-making, and executive orders. Often particular environmental goals are embodied in legislation such as the Clean Air Act or Safe Drinking Water Act, in a myriad of state and local laws on environmental quality, or in international treaties such as the Montreal Protocol on Substances that Deplete the Ozone Layer. Increasingly, environmental goals are also stated by organizations outside the public sector, such as Monsanto's 1990 pledge to reduce its toxic air emissions by 90% (Smart, 1992).
The nation pursues a variety of activities to achieve its environmental policy goals. Research and development goals are one important component of such government actions, and represent the third level in our hierarchy. Research and development goals answer the question: "what information and technology does the federal government and other actors need to determine and carry out the environmental policies?" (Consistent with our broader definition of Level 3 goals, this phrasing expands on that of the CENR guidance which focuses on scientific information only.) For example, research and development goals include developing improved models to predict the long-term changes in the climate and publishing an ecosystem map for the United States at a scale which allows local decision-makers to incorporate spatially explicit environmental factors into urban planning and resource management decisions [NSTC, 1995].

It is important to recognize, however, that research and development is only one of many possible government actions which can play a part in achieving environmental goals. For instance, the federal government generally uses a wide variety of regulatory, land use policies, and procurement policies to enhance the nation's environmental quality. Research and development goals focus on generating the information that public and private sector decision-makers will need in the long- and short-term to more effectively improve environmental quality while also pursuing other societal goals, such as economic growth. Research and development goals are determined by the managers of R&D programs within the federal agencies and by recommendations of professional societies. Congressional committees may also express opinions about research and development goals. For instance, the CENR strategic plan [NSTC, 1995] provides examples of such research and development goals: i) an improved understanding of the socioeconomic dimensions of environmental change in order to better discern the underlying human influences on the environment and the potential responses of society to change, ii) improved science policy tools so that policy-makers can make more informed decisions on complex environmental and social issues, iii) improved ecosystem research to promote efficient use of natural resources while sustaining ecosystem integrity for future generations, and iv)
improved environmental technology so that US firms can better protect the environment while stimulating economic growth and capturing emerging global markets. Note that these goals represent a broad range of different types of research and development activities.

Figure 3 shows an example of one environmental policy goal and its supporting research and development goals. For the purposes of this example, we have adapted part of the SBED subcommittee's draft implementation plan into the structure recommended by our template. This part of the draft addressed the scientific area of understanding the system structure and function of ecosystems. An important environmental policy (level 2) goal relevant to SBED is to protect the environment for present and future generations while enabling development to proceed in a sustainable manner. In order to achieve this goal, policy-makers must answer the question how will the environment behave in response to various actions? An R&D (level 3) goal to gather information necessary to answer this question is: develop scientifically sound, user friendly information and validated analytical models to enable exploration of possible consequences of alternative management and policy decisions. Three program (level 4) goals necessary to achieve the research and development goals are shown in Figure 3. These goals answer the questions: what is the current status of ecosystems and biodiversity, how do ecosystems behave, and how will ecosystems behave under future scenarios and policy choices?

It is important to note that individual research and development goals can meet several different higher level goals. For instance, new systems for collecting environmental data could help provide a better understanding of global warming as well as providing farmers with better long-range weather predictions. Environmental technologies could help achieve the national goal of improved environmental quality as well as job creation and economic growth. We will return to this theme in Section 3.2 on performance measures for program balance.

Level 4 in our hierarchy are the objectives for individual research and development programs. Program goals address the key questions and
understanding which is required to advance the nation's research and development goals. While different federal agencies have programs of different sizes, these goals may be thought of as questions which could be answered by a series of requests for proposals (RFPs). As shown in Figure 3, for example, three Level 4 program goals are necessary to address the Level 3 goal of developing validated analytic models: understanding the current situation, developing the models, and using these models to manage ecosystems. Generally these goals are the focus of the program managers within government funding agencies, project managers conducting intramural research, and individual researchers who will write grants and plan their own activities based on these Level 4 goals.

When developing these level 4 goals, CENR subcommittees should carefully differentiate among the types of R&D they sponsor. Often this is not done because of a natural tendency to lump different types of R&D together, as we noticed in our work with one of the subcommittees. Also, it can be hard to clearly distinguish the different types of R&D since they are often very interrelated tasks. However, it is worth taking the time and effort to think through and differentiate the different types of R&D goals at Level 4. This distinct classification is necessary to facilitate the choice of performance measures, as described in the next section. Some types of R&D are more amenable to quantitative outcome measures than others. By distinguishing the different categories of their research and development, the subcommittees can chose the most effective measures for each type.

For instance, we have identified four categories of CENR research and development --science, technology, data collection and management, and assessment -- which cover a wide range of CENR activities. Science research aims at improving our understanding of the environment, the ways in which human activities affect the environment, and the ways in changes in environmental quality affect society. Technology R&D aims to improve our effective options for reducing the impacts of human activities on the environment, reducing past environmental degradation, and monitoring the environment [NSTC, 1994]. Data collection and management aims to assemble, process, store, manage and disseminate the environmental and societal data and information which researchers
and decision-makers need to pursue their goals. **Assessment** aims to
synthesize, analyze, and translate complex ecological, socioeconomic, and
health data and scientific understanding to help public and private policy
decision-makers pursue their goals. Assessment includes techniques
such as integrated, risk, and cost/benefit assessments,\(^1\) and other similar
analysis efforts to help support policy decisions.

Each of these four R&D categories has different performance
measures most appropriate to each. For instance, measures of economic
impact appropriate to technology R&D and measures of policy impact
appropriate to assessment are not appropriate for scientific research. The
reader should note, however, that not all CENR subcommittees will
necessarily have research in each of these categories and some
subcommittees may find it advantageous to develop additional categories.

\(^1\) For specific definitions of integrated, risk, and cost/benefit assessment see CENR co-
chair's memo on "Data Collection and Formats for FY 1996 & Environment and
Natural Resources Terms of Reference", June 29, 1994.
3. PERFORMANCE MEASURES FOR CENR

Establishing a hierarchy of goals is an important first step in developing performance measures for CENR research and development activities. However, choosing measures which demonstrate how well R&D programs address these goals remains difficult. The challenge facing CENR, and most federal sponsors of research and development, is to develop performance measures responsive to Congressional and public calls for accountability of public funds, which at the same time are effective tools in managing research and do not damage the underlying research enterprise.

The template presented here addresses this challenge by distinguishing two broad applications of performance measures. These are measures for individual programs, which address how well individual programs at Level 4 in the hierarchy contribute to higher level goals, and measures of program balance, which address how well groups of research and development activities taken as a whole contribute to higher level goals.

Performance measures for individual programs are the application envisioned by the GPRA and NPR. In addition, such measures are important for the program managers and researchers themselves directly concerned with the performance of individual research and development programs. Some types of CENR research and development can have more effective performance measures than others. As we will see, the template is designed to help choose the most appropriate and effective type of individual performance measures for each type of CENR R&D.

It is unlikely, however, that any set of performance measures for individual research and development programs can truly provide insight into how well CENR R&D supports environmental policy and higher level national goals. It is often very hard to predict the outcome of development efforts in relationship to policy goals. Furthermore, research is fundamentally unpredictable, and no matter how well we measure
today's research activity, we can not anticipate the value of tomorrow's outcomes with any precision. Nonetheless, we may be able to develop useful measures of the contribution of R&D by focusing on those features of R&D activities associated with successful outcomes. Even if we can't foresee the outcomes of individual lines of inquiry, we may be able to describe the characteristics of an overall research and development effort which is likely to achieve its goals.

The measures of program balance described in this template adopt such a focus and view CENR R&D as a portfolio of activities. A portfolio aims to spread risks over a number of investments. While any single investment may have a high chance of failing to meet its goals, the portfolio as a whole can be highly likely to succeed. The portfolio concept is common among R&D managers in the private sector, who often evaluate their R&D activities as a mix of investments distributed among high risk, high pay off and low risk, moderate payoff programs covering the full spectrum of the firms' potential business activities.

While government R&D managers do not always use the portfolio concept explicitly, they often describe guiding principles for their R&D activities which are characteristics of the activity as a whole rather than the individual parts. For instance, CENR's strategic plan offers a list of principles which calls for environmental research and development to be policy-relevant and not policy-driven, anticipatory to prevent and/or mitigate serious environmental threats, and completely integrating the social and natural sciences [NSTC, 1995]. CENR can use measures of program balance to evaluate the extent to which their programs meet such criteria. In addition, balance measures are also useful because they explicitly recognize that multiple research and development programs are often necessary to achieve particular higher level goals.

As suggested in Figure 2, CENR subcommittees can develop effective performance measures by employing a mix of measures for individual

---

2 Some federal R&D managers do explicitly use portfolio concepts. For instance, NSF's Science and Technology Centers use evaluations which consider their research activities as portfolios [Cozzens, 1995].
programs and measures of program balance. After first articulating a hierarchy of goals, as described in Section 2, CENR subcommittees should choose performance measures for their individual Level 4 programs as described in Section 3.1 below. These measures will help CENR managers and those with oversight over CENR evaluate how efficiently individual programs are run and how well each contributes to its (level 4) program, and perhaps (level 3) research and development, goals. The subcommittees should then choose performance measures of program balance, as described in Section 3.2, to evaluate how well groups of CENR programs contribute to higher level goals. This mix of performance measures will help CENR to address the growing demands for public accountability without damaging the vitality and utility of their research and development activities.

3.1 Measures for Individual Programs

Table 2 shows twenty-three techniques which CENR subcommittees could potentially use as performance measures for their individual research and development programs. We intend Table 2 as a menu of options from which the CENR subcommittees can pick the combination of measures most appropriate for their own programs. In the discussions which follow, our brief descriptions provide basic information on each of these measures to help readers select those which appear promising. The reader can then refer to the specific literature cited herein to obtain a more complete description of any measure they would like to consider using in their strategic and implementation plans.

None of the measures in Table 2 are a panacea. Most of the measures which are most quantitative and/or the simplest to use also provide the least reliable information on a program's performance and its contribution to higher level goals. Those measures which offer the most information on a program's contribution to goals often require significant effort and several years of gathering data to use effectively. Many of the most intriguing measures of program outcomes require time consuming effort to identify and contract the stakeholders -- such as the researchers, public policy-makers, private sector decision-makers, and environmental
groups -- who use or could use the information generated by the program. In practice, CENR will likely use a mix of measures designed to balance each others' weaknesses. Our purpose here is to help CENR better understand the available albeit imperfect options, and to place these options within a conceptual framework where they can be used most effectively.

The performance measure options in Table 2 are adapted from those used by government managers of fundamental science and industry managers of applied research and development. Our source for the former is the CTI project on the metrics of fundamental science, conducted for the OSTP Associate Director for Science and the NSTC Committee on Fundamental Science. This project was designed to collect and summarize the knowledge and experience of scholars and government research managers concerned with metrics for fundamental science. This project consisted of three parts: i) a detailed literature review [Popper, 1995; Wagner, 1995; Cozzens, 1995; Melkers, in draft]; ii) a one-day workshop in November 1994, which brought together practitioners, academic theorists, and policy analysts to discuss available tools and methods and their applicability to measuring and assessing science [Wagner and Flanagan, 1995]; and iii) a colloquium held in May 1995 of 150 experts, including scientists, practitioners, stakeholders, and analysts, to discuss the new emphasis on programmatic accountability as it may affect fundamental science, and to develop issues and options for building assessment strategies that will fully represent the contribution of science to national well-being.

Our source for the metrics used by industry R&D managers is the Menu of Metrics developed by the Industrial Research Institute (IRI) (see Appendix B). The IRI is an organization of more than 260 major industrial firms with an interest in the effective management of research and development to stimulate technological innovation. The IRI member companies represent about 80 percent of the nation's privately funded research effort and approximately 30% of gross domestic product. The

---

3 One of the authors of this report, Darius Sankey, is a member of the IRI.
list of IRI metrics was compiled by a group of research directors in order to develop a software program for various stakeholders to evaluate the research investments of a corporation. The emphasis was placed on business and R&D managers who use and create technology. A subset of the IRI metrics have been successfully applied in several corporations, including AT&T Bell Laboratories, Eastman Chemical, Dupont, 3M, Hughes Aircraft Research Labs, General Electric, and IBM. However, some of the IRI metrics are conceptual and have not been fully tested. In Appendix B, we provide a full list and description of the IRI metrics. Elsewhere in this report, we will focus only on those we believe may be useful to CENR.

CENR can use ten of the measures in Table 2 to directly evaluate the extent to which an individual program contributes to its goals or to a higher level goal. Such measures, which assess how well the program contributes to higher level goals, we call **Performance towards Goals** measures. These measures focus on program outcomes rather than program inputs, and as such, address CENR's primary interest in performance measures. Examples of Performance towards Goals measures include peer review of the relevance of a research program to national goals and estimates of the social return on the investment in a program. Such measures are not only useful for program managers, they are of great interest to policy-makers such as OMB and Congress who can use them to better understand how they can allocate available resources to achieve broad national goals.

Wherever possible, CENR subcommittees ought to chose Performance towards Goals measures in their strategic and implementation plans. In many cases, however, particularly for fundamental science programs, adequate Performance Towards Goals measures are not available. In these cases, CENR subcommittees can use measures which assess how well the program is being carried out, which we call **Process Evaluation** measures. Process Evaluation measures evaluate the efficiency of a program, including its productivity and managerial performance. There are thirteen Process Evaluation measures in Table 2. While an improvement over input measures
because they measure how well a program's inputs are used, Process Evaluation measures fall short of true outcome measures because they do not directly link the measured activities and the goals of the program. Examples of process measures include the number of peer-reviewed papers generated by a scientific research program and the number of programmatic milestones achieved. Process evaluations are often very useful for program managers, who can use them to make their programs run more efficiently. While higher level policy-makers may find process measures useful in determining which programs are well and poorly run, they are only indirectly useful in allocating resources among programs. A poorly run program may in fact be much more important to achieving key national goals than a less relevant, but well-managed activity.

In addition to the distinction between Performance Towards Goals and Process Evaluation measures, Table 2 also distinguishes between ex-ante (before the completion of a program) and ex-post (after completion of a program) measures. Ex-ante measures are more desirable, since they provide the possibility of modifying, or even ending, a program if needed. However, ex-ante measures are often less reliable than ex-post measures. Many of the former have been derived by observing process characteristics of programs which were found, ex-post, to be successful, and assuming that, if these characteristics are observed in an ongoing program, that program is likely to also be successful. In addition, it may take many years before ex-post evaluations can be conducted on R&D programs, which decreases their utility as a tool for research management and accountability.

Ex-ante Performance Towards Goals measures are the type of performance measure for individual programs which best satisfies CENR's concerns. Unfortunately, as is shown in Table 2, we have only been able to identify three candidate techniques in this category applicable to CENR programs from the many types of metrics used by the government to manage its fundamental science programs and by industry to manage its applied research. The fundamental science measure applicable to the most CENR programs is the well-known
technique of peer review, in this case applied to the relevance to program goals. As shown in Table 2, peer reviews for program relevance can be used for all four types of CENR research and development programs.

Peer review, of course, is the most common and most fundamental type of evaluation method for research programs. Peer reviews consist of ex-ante evaluations of research project proposals by researchers with specialized expertise in the project area. It has become increasingly common in recent years for peer reviews in many government agencies to use peer review for judgments about the relevance of proposed projects to program goals in addition to judgments about the project's scientific merit [Ziman, 1994] (We discuss this second type of use below). The well-known strength of peer review is that it is perhaps the only method of scientific evaluation that has any chance of predicting the ultimate quality and usefulness of a proposed research project. By definition, high quality research projects are at the frontiers of human knowledge and only a small number of experts will likely to be in a position to competently appreciate the potential of a specialized research proposal. Peer review also has well-known problems [GAO, 1994]. It demands extensive time from researchers writing and evaluating proposals, which could otherwise be spent producing useful research. Peer review raises questions of equity among minorities and institutions which have not traditionally been part of the research system. In addition, peer review can be risk adverse and can discourage interdisciplinary work and work in new fields. This is a particular concern of CENR, since environmental and natural resources R&D is widely recognized as an area where interdisciplinary work is required. Finally, peer review for program relevance may be hindered by an inability to find suitable experts who sufficiently understand both the proposed research and its potential applications. Indeed, not even the researchers proposing the project may have any good idea of what its applications will be.

The IRI Menu of Metrics offers thirty-three performance measures, most of which are potential applicable to some or all types of CENR research and development programs. Nonetheless, CENR can only use two IRI metrics to satisfy its interest in ex-ante Performance Towards
Goals measures. This is probably not surprising given the differences in the types of R&D conducted by firms and the government, and the different goals for which their research is pursued. We expect CENR can adopt two ex-ante Performance Towards Goals measures: programs with user approval (IRI Metric 10) and the percentage of outside funding supporting the program (IRI Metric 12). The former, for a business, measures the fraction of projects in the total R&D portfolio with explicit business unit and/or corporate business management sign-off. CENR could use this measure as the fraction of programs or projects with have been vetted and/or approved by stakeholders likely to use the results of the research. Such stakeholders might include policy-makers in federal, state, and local governments, leaders in the business community, and environmental groups. CENR may, however, find it difficult and/or costly to identify and adequately survey the stakeholders for any given program. The latter measure, for a business, measures the fraction of the R&D budget contributed by business units which are potential users of the research. CENR could use this measure as the fraction of the R&D budget of certain programs contributed by potential private or public sector users. For instance, many government technology R&D programs, such as EPA's Environmental Technology Initiative and the Commerce Department's Advanced Technology Program, encourage cost sharing among recipients of government funds, and CENR subcommittees could report the number of their relevant programs which employ this method.

Table 2 also identifies seven metrics which CENR sub-committees can use for ex-post evaluations of the performance of individual programs towards higher level goals. Program evaluations [Cozzens, 1994] are an examination of the quality and relevance of the implementation and results of ongoing and completed research programs. These evaluations are a powerful technique used by many government agencies. Though both are carried out by panels of specialists, program evaluations differ from peer reviews in two important ways. First, program evaluation applies to programs rather than projects. The focus is the combination of

---

4 Program evaluations often include stakeholders who use the research in the evaluation panels as well as experts in conducting the research.
the program's goals and objectives, the portfolio of activities it supports, and the results these activities produce together. Second, a program evaluation gathers information about actual performance rather than projecting the performance promised in a proposal. In addressing both the quality and relevance of a set of activities, program evaluations address precisely the correct issues for performance measures. They must be used sparingly, however, since they can be expensive and time consuming to conduct. In addition, program evaluations do not produce quantitative measures of performance.

Several government agencies have conducted retrospective studies to produce quantitative measures of the importance of research and development programs to particular national goals. Among the most well-known, Project Hindsight attempted to identify the contributions of Department of Defense research programs to improved cost-effectiveness in weapons systems and benefits to national security [Department of Defense, 1969] and TRACES (Technology in Retrospect and Critical Events in Science) was a large-scale, retrospective look at the accomplishments of fundamental science [Illinois Institute, 1968]. While such retrospective studies provide very valuable information, they are very expensive and generally produce the best information on programs which are a decade or more old. Accomplishment books, are a much more timely, but much less quantitative, technique. Accomplishment books are documents which present descriptions of selected scientific accomplishments resulting from selected programs, and then describe the contributions these accomplishments have made to national goals. Most agencies produce accomplishment books for a broad audience, including funding sources such as Congress and the Executive Branch.

Table 2 provides several metrics which can provide quantitative performance measures for some types of CENR programs in the areas of data, technology, and assessment programs, but which are probably not relevant to science programs. User ratings of benefits are a technique adapted from the IRI's customer ratings of product technology benefits (Metric 15). For a business, customer ratings are a numerical ranking of a firm's product by a given customer divided by that customer's ranking
of the best competitive product. The ratio can be averages across customers and/or market segments using the product to obtain an average ratio value. CENR can conduct surveys of users' ratings of the usefulness of CENR research and development (and perhaps other government R&D programs) and compare the numerical ratings of particular CENR programs to the best rated programs in the survey. Users of CENR research and development include policy-makers, private sector firms planning to commercialize or use technologies funded by CENR technology programs, and researchers using data collected in CENR data programs. Some government agencies, such as the Office of Naval Research, employ user ratings as part of their evaluation procedures. One danger of such ratings is the potential upward bias given by users aware of their role in determining the future of a government service they are provided for free.

**Impacts of research**, adapted from IRI’s economic evaluation rating (Metric 15), are monetary measures of the benefits of a research and development program. R&D programs which are now part of CENR have estimated such impacts in the past. For instance, CENR R&D has produced warnings which allow farmers to avoid storm damage to crops and technologies which have increased energy efficiency. **Timeliness of research** (adapted from IRI Metrics 16 and 19) measures whether a research program has supplied policy-relevant information to decision-makers early enough to affect the decision-making process. This measure is primarily useful for CENR assessment programs.

Table 2 also identifies six techniques which can be used by the CENR sub-committees for ex-ante Process Evaluations individual programs. **Peer review** evaluations of the quality of proposed projects and the track records of proposed investigators is of course the fundamental such measure and is used by most government agencies funding research as well as by most members of IRI (Metric 23).

Table 2 offers CENR two measures of the quality of program staff, their management, and their working environment. **Quality of personnel** (IRI Metric 18) can be measured by a interval rating scale by the users of a
program's research, by external recognition such as number of external awards or invited lectures, or number of published works or patents. **Employee morale** (IRI Metric 29) is a quantitative rating of key aspects of employee satisfaction and morale shown by direct employee survey. These two measures are relevant to all types of CENR research. **Use of milestones** (IRI Metric 11) is the fraction of projects in the program going through a defined project management system with defined milestones. The success of some types of R&D, particularly assessment, depends critically on the sharing of information among researchers in different disciplines and different institutions. The extent of such linkages can be measured by adapting the measure **Use of Cross-Functional Teams** (IRI Metric 14), which is the fraction of projects in the research program with specific cross-functional team assignments.

Finally, Table 2 identifies seven techniques for ex-post Process Evaluation of CENR programs. In the United States, the use of **technical reviews** varies from a very informal assessment process to highly retrospective quality control mechanisms. The Office of Energy Research at the Department of Energy, for instance, conducts hundreds of highly structured project evaluations each year in which expert reviewers rate projects according to standard categories. These quantitative ratings are then compared across projects to identify those which need improvement [Cozzens, 1995]. Such technical reviews differ from peer review because the experts assess project accomplishments rather than project proposals. For businesses, **Customer Rating of Technical Capability** (IRI Metric 20) compares the average customer rating of the overall technical capability of the firm in providing technical service and/or new product innovations to ratings of relevant competitors for benchmarking purposes. CENR's 'customers' for such reviews would be the potential users of the research, for instance, policy-makers who might use the results of a CENR assessment, firms which might commercialize new technologies based on CENR technology R&D, and scientists who might use information from CENR data collection. Technical reviews produce useful project management information, but they are not a performance measure panacea since they are expensive and can suffer from ratings
inflation when used to allocate resources among different types of programs.

**Publication counts** are the most widely used metric of knowledge production in science, used in applications ranging from promotion decisions to assessments of the productivity of different nations' scientific establishments [Cozzens, 1995]. CENR will likely find publication counts are most appropriate for its scientific research. Nonetheless, comparisons of dissimilar programs is complicated by the different publications habits in different disciplines and different types of institutions. For instance, chemists usually write many short papers while geologists publish occasional massive reports of their work. Academics regard publications as their primary output, while engineers can spend much of their time producing new technology whose details are shielded from public view. Because publication counts often only measure the quantity of research, many agencies also use **citation counts** as a measure of the quality of scientific research. Citation counts are also best compared between similar types of research programs since different fields can have different citing practices. Practitioners should also note the time lags involved. Citations tend to peak two or three years after publication of an article.

The management of those R&D programs for which milestones are appropriate can be evaluated by the number of **milestones achieved**. The measure Development Pipeline Milestones Achieved (IRI Metric 25) suggests assessing the percent of project milestones achieved within a certain margin of the projected achievement date, plotted as histograms to reveal actual performance, showing percent that beat the target as well as those that miss it. Those CENR programs in technology and assessment which produce services for specific users can also be evaluated by Customer Contact Time (IRI Metric 26), the average hours per researcher spent in direct contact with users of the programs' research.

Table 2 also provides CENR two useful measures of the impact of CENR technology programs. **Technology transfer** to Manufacturing (IRI Metric 13), is a semi-quantitative assessment on an interval scale (e.g. 1 to
5) of the effectiveness of the transfer with ratings obtained from both research and manufacturing sides of the interface. CENR can adapt this measure to include any user community of CENR technology programs. Number and Quality of Patents (IRI Metric 21) can also be adapted for CENR use. IRI recommends measuring the percent of useful patents, value ratio of patents, the retention percent, and the cost of invention.

3.2 Performance Measures for Program Balance

It is clear from Table 2 that measures of individual programs are insufficient to manage CENR investments in environmental and natural resource research. Only three ex-ante measures are available to link individual programs to higher level national goals. Only peer review is applicable to many CENR programs, but it is not quantitative. None of the three is particularly useful for comparing across programs.

This inadequate supply of performance measures for individual programs is a key example of CENR's need for performance measures of program balance. This section offers eight such balance measures for CENR use. All are adapted from the IRI metrics, which gives significant emphasis to the management of portfolios of research. The IRI divides its metrics into five major categories, four of which are of primary interest to CENR. The first two categories of interest, practice of R&D processes and integration of R&D with the business, provide CENR with the IRI-derived performance measures for individual programs discussed in Section 3.1. Metrics in the former category relate to the efficiency with which R&D is conducted and provide CENR with process evaluation measures (e.g. quality of personnel). Metrics in the later category indicate the degree of integration between the groups conducting research and the groups using research, the commitment of a firm's product divisions to the research processes and programs, and the ability of the organization to exploit technology. Such metrics provide CENR

---

5 Some government agencies such as NSF [Cozzens, 1995] and NIST [NIST, 1994] also use balance measures for their research programs.
with performance towards goals measures (e.g. percentage of programs with outside funding).

The second set of IRI categories, **portfolio management** and **value creation**, provide CENR with the balance measures discussed below. Measures in the former category communicate information to high-level decision-makers about the effectiveness the total research and development program and allow the entire program to be structured to improve its overall benefit to the corporation. CENR can use such portfolio measures as a benchmark for the proper allocation of funds across a diverse set of activities. Measures in the latter category directly demonstrate the value of research and development activities to the positioning, profitability, and growth of the corporation of shareholder value. While CENR can occasionally apply a value creation metric as a performance measure for an individual programs (e.g. user's ratings of benefits in Table 2), the committee will in generally find value creation measures most useful for evaluating the contribution groups of its R&D programs make to achieving environmental policy and national goals.

Table 3 shows how the IRI ranks the importance of these categories of metrics to the different corporate stakeholders concerned with the research investments. Note that value creation and portfolio assessment are the two most important types of measures for a firms' highest level external stakeholders, the board of directors and the financial community. The analogous highest-level stakeholders for CENR are the Congress, OMB, and the public. Table 3 offers insights on how CENR can use different types of performance measures to address the concerns of different audiences, and in particular, which measures can articulate to high-level policy-makers the need for a well-balanced program of high-quality CENR research, which spans a range from basic science to technology development, and addresses both the near-term and long-term needs of policy-makers and the public.

Table 4 shows six measures of program balance which CENR can use to evaluate how well groups of its research and development programs satisfy higher level goals. As with performance measures for
individual programs, the Table distinguishes between ex-ante and ex-post measures. As before, ex-ante evaluations will best satisfy CENR's interest in performance measures. Table 4 provides three ex-ante measures of program balance. The most quantitative of the metrics is **projected value of the research** (adapted from IRI Metric 2). A profit-making firm can estimate the projected value of its research portfolio by calculating the fraction of future sales by year, and net income from these sales, projected from projects in the research pipeline, incorporating the probability of that each project attains its objectives. CENR will have to use this measure somewhat differently. For many technology projects, CENR can attempt to estimate the cost savings that will accrue to society from widespread use of the fruits of the research. For example, five Department of Energy laboratories estimated the economic benefits society would gain from its investments in renewable energy technology research [INEL, 1990]. This metric could be similarly applicable to many CENR technology research and development activities.

For CENR R&D aimed at improving the decisions society makes about environmental policy, the metric would need to address the value of the research to improved decision-making. The decision theory literature defines the project value of improved information (research) as the difference in costs to society of making its decisions with and without the information that the R&D programs are expected to generate. The literature on climate change, for example, already contains a number of such estimates of the value of improved information about different aspects of the climate change problem, generated with large-scale computer models [Manne and Richels, 1992; Nordhaus, 1994]. Other scholars estimate the value of improved information about environmental and natural resources research through expert elicitation, a formal survey technique of experts in the relevant scientific fields [Morgan and Henrion, 1990]. In many cases, however, such techniques are still too primitive and costly to be effective tools for CENR. Improving the ability to make such estimates of the value of information is one of the topics likely to be addressed under the research goal of improved science policy tools, one of the priority areas identified for FY96 in the Gibbons-Panetta memo.
Table 4 also offers two less quantitative ex-ante measures of program balance. **Strategic alignment** (IRI Metric 7), a portfolio management measure, evaluates the extent to which a research portfolio addresses the goals to which it is addressed. IRI suggests that profit-making firms enumerate the fraction of the firm’s goals that are addressed by the research portfolio and enumerate the fraction of the total research portfolio which is consistent with corporate goals, and where, applicable, each major business unit in the firm. These goals and business units served by each R&D program should be explicitly identified. Using the hierarchy of national goals in Figure 3 to replace the IRI corporate and business unit goals, CENR should be able the use the strategic alignment measure for most of its R&D activities. For example, CENR could total funds budgeted for the R&D (level 4) programs which support each of its environmental policy and national goals. In addition, CENR could list the number of higher level goals supported by each R&D program. An application of the strategic alignment measure should also provide CENR information on gaps and redundancies in its overall R&D effort.

**Distribution of Research Investments** is derived from a private sector measure of similar name. IRI’s Distribution of Technology Investment (Metric 7), a portfolio management measure, suggests that profit-making firms analyze the fraction of their total research investment along various dimensions which are important to decision-making within the firm and communications among stakeholders. Such dimensions might include: categorization of risk and reward, by product line or business unit, distribution according to time to commercialization, or current technology vs. technology new to the company. CENR should be able to develop a set of suitable dimensions over which to examine their R&D investments. Example might include: research to address current policy needs vs. research to anticipate future needs; research on the physical environment vs. research on the socioeconomic dimensions of environmental protection; research directed at problems which are primarily a federal responsibility vs. problems which are primarily the responsibility of state and local government; and research directed at providing a better understanding of potential problems vs. research directed at generating potential solutions to problems. While distribution of investments has the
potential to be a very useful measure, CENR will need to conduct systematic reviews (benchmarking exercises) of previous R&D activities and research in other organizations to gain insight into the proper dimensions to use and numerical estimates of fraction of investments along each dimension which characterize a successful R&D portfolio.

Table 4 also provides three ex-post measures of program balance. **Financial returns** (IRI Metric 1) is the fundamental value creation metric for a profit-making firm and is a retrospective assessment of the sales, cost savings, and return on investment from a firm's R&D portfolio. Similarly to the situation with the ex-ante measure projected value of research, CENR can adapt financial returns to its own programs in a limited number of cases. CENR can estimate social benefits such as cost savings or improved environmental quality for some of its technology R&D. For instance, DOE has conducted studies which attribute large economic returns to federal investment in technology for energy efficiency [Rosenberg, 1991]. In principle, retrospective studies of the value of information would also be possible for much of the CENR R&D portfolio.

CENR can also measure the benefit of its R&D portfolio by systematically surveying those who used or should have used the results of the research. This **users' ratings of benefits**, a value creation measure, is derived from IRI Metric 15, which also appeared as an ex-post performance measure towards Goals measure for individual programs. CENR can also employ users' ratings for to groups of programs using the same survey techniques as discussed in Section 3.1. For example, a CENR subcommittee could conduct surveys to learn which types of R&D its stakeholders find most valuable. While such evaluations can provide useful information for allocating resources, they must be conducted and interpreted carefully since the results can be strongly influenced by the stakeholders included in the survey.

Finally, Table 4 offers number of **ways research exploited** (IRI Metric 9), as a Portfolio Management measure in which a firm assesses the number of different product types or business segments utilizing or planning to utilize a given technical asset. CENR could adopt this
measure as an assessment of the fraction of the policy-decisions influenced by relevant areas of CENR research. For instance, CENR could examine the extent to which information generated by CENR R&D has an important influence on the decisions made by environmental regulators at the national, state, and local levels.
4. USING THE TEMPLATE

The CTI template played only a small role in helping CENR and its subcommittees develop their strategic plans. Its utility was limited because our work with CENR began after the committee was well into its FY96 planning process. Nonetheless, the CENR Strategic Planning Document, issued in March 1995, employs many of the concepts discussed in this report. The CENR document provides an overview of CENR strategic planning and coordination of federal environmental research and development and twelve summaries of the strategic plans of each of the CENR subcommittees. These plans use both a hierarchy of goals and several types of performance measures.

The document gives significant attention to developing a hierarchy of goals for CENR research and development. The first chapter reviews the broad national goals supported by federal research and development and the environmental policy goals supported by CENR research and development, providing a list similar to that shown on page 10 of this report. Each of the strategic plans for individual subcommittees describes the highest priority research and development goals needed to support the environmental policy goals most germane to each particular subcommittee. For instance, the Subcommittee on Global Change Research lists the Framework Convention on Climate Change, the U.S. Climate Change Action Plan, the Montreal Protocol on Substances that Deplete the Ozone Layer, and the U.S. 1990 Clean Air Act Amendments as among the official documents describing the policy goals for their subcommittee's research. The subcommittee's strategic plan lists climate change and greenhouse effect research as one of the high priority research areas needed to meet these policy goals. The subcommittee offers four goals for this research: 1) observe and document human influences on long-term climate change; 2) improve the capability of climate system models to represent relevant climate processes and better predict global and regional changes in the climate; 3) predict the consequences of climate change on natural and socioeconomic systems, and 4) support integrated assessments of climate change, including research on the feasibility and effectiveness of response options.
The CENR plan also offers a two types of performance measures for individual programs for evaluating CENR research and development. Each of the CENR subcommittees' strategic plans provides programmatic milestones for the next three years of research (an ex-post, process evaluation measure), and brief accomplishment books by presenting two examples of noteworthy impacts from previous R&D activities (an ex-post, performance towards goals measure). For instance, the GCRP offers the milestone that by 1998, their research will provide regular forecasts of the timing and distribution of extreme climate events such as flooding and droughts related to seasonal or interannual climate variability (such as the El Nino-Southern Oscillation). GCRP also points to the understanding of the Antarctic 'ozone hole' as crucial policy-relevant information provided by their research. The CENR strategic plans do not offer any performance measures for program balance. However, the first chapter of the document provides a number of characteristics and principles for effective CENR research which could provide initial criteria to develop such balance measures.

CENR is likely to expand and further refine its use of performance measures in the future. The template can help in this process. To that end, it is useful to review how this template can be used by CENR. This template was designed for CENR as a 'how to' guide and reference manual for developing performance measures for their research and development plans. The structure of the template reflects the steps CENR can go through when developing such measures. Although we present these as a series of steps, they are more akin to an iterative procedure. To be effective, performance measures must be designed into strategic and implementation plans from the beginning, and both must be readjusted so they fit together. These steps are:

1) The subcommittees should review the environmental policy goals (level 2) and the research and development goals (level 3) which support them, as laid out in their strategic plans. The arguments justifying their research and development priorities, called for by the strategic plan guidance, will likely draw upon arguments
similar to the performance measures for program balance shown in Table 4.

2) The subcommittees should review the program goals (level 4) they will include in their implementation plans. They should divide these goals in categories, such as the science, technology, data collection and management, and assessment categories we use here, or other groupings as the subcommittee finds appropriate.

3) The subcommittees can then choose performance measures for the individual programs from the list in Table 4. Ex-ante ‘performance towards goals’ measures should be used where possible, with other types used as necessary. In practice, each program will likely use a mix of measures of different types.

4) The subcommittees can chose measures for program balance, from the list in Table 4, to help justify the choice and mix of different types of program goals (level 4), and to help evaluate the contribution of these programs to research and development (level 3) and environmental policy (level 2) goals.

5) Use the performance measures to monitor the progress of CENR research and development activities, and to modify these activities as necessary as more is learned about the efficiency of various programs, the promise of the research they support, and their usefulness to CENR's and the nation's goals.

Our work with the CENR subcommittees during the last budget cycle suggests several lessons which CENR should keep in mind while using the template in future years. First, it is important that CENR make performance measures part of its planning processes from the very beginning. While developing the goals it will use in its strategic and implementation plans, CENR must choose and expressed these goals in such a way as to facilitate the use of performance measures. For instance, Level 4 program goals should wherever possible carefully differentiate among different types of CENR research and development, so
that the most effective set of performance measures for individual programs can be chosen for each. As we saw in Table 2, some types of CENR R&D, such as technology, have far more options for quantitative performance measures, than do other types of CENR R&D, such as fundamental science. The goals at different levels of the hierarchy and the links between them should also be phrased clearly and concisely in order to facilitate the use of performance measures for program balance. Such measures work best when applied to programs which make significant contributions to higher level goals, and where the links between a program and the multiple goals to which it could contribute are made clear. In addition, CENR managers and planners need to know what performance measures will be used so that they have a fair opportunity to direct their programs in such a way that they will receive positive evaluations. On one occasion in our work with the CENR, a subcommittee declined to use a strategic alignment measure of program balance which used budget data over several years to show trends in the spending on R&D programs intended to support different R&D goals. The trends seemed to be contrary to the stated priorities of the subcommittee.

Second, CENR will have to gather significant amounts of data in order to make use of many of the potentially most effective performance measures. This data can include detailed budget information for CENR and other programs, surveys of user satisfaction with CENR R&D, and the economics and other data needed to estimate the impacts of CENR R&D. In many cases CENR will need to establish trend data over several years, and/or gather data on non-CENR programs, in order to show improvement over time and establish benchmarks for judging how well different CENR efforts compare to comparable R&D activities. For instance, the metric Distribution of Research Investments can be a powerful measure of program balance. It requires, however, that CENR divide budget data on its R&D portfolio into categories such as near-term vs. far-term and fundamental science vs. technology research. Although new tools such as the RaDiUS database developed by CTI may now make it
feasible to gather such data, CENR will still have to devote sufficient effort to develop the appropriate categories and continue these efforts over several years to provide a set of time series for meaningful comparisons. Similarly, measures such as Programs with User Approval, a measure for individual programs which directly addresses the CENR's concerns, would require a process whereby potential users of CENR research have an opportunity to vet CENR programs during the planning process. CENR would have to maintain the effort over several years and among different subcommittees in order to develop a body of data so that performance of programs could be meaningfully compared over time and to one another.

Third, CENR should use a small number of carefully balanced performance measures for evaluating its programs. It is not yet clear to us how many is too many. However, CENR should think in terms defining a set of measures which are common among all its subcommittees, as well as allowing individual subcommittees to use a small number of measures most applicable for their research. The measures should be balanced to compensate for each others' weaknesses and exploit different strengths. For instance, CENR might complement several inexpensive measures designed to produce annual indicators of program progress with infrequent and more expensive measures designed to evaluate contributions to higher level goals. CENR should probably experiment with a number of different measures for several years before deciding which it finds most useful.

Performance measures will remain a significant challenge for CENR and other federal research and development programs. There are significant dangers to incomplete, excessive, or poorly conceived performance measures. However, the demands for accountability in

---

6 RaDiUS (Research and Development in the United States) is the first comprehensive, real-time accounting of federal R&D activities and spending. RaDiUS allows users to track federal R&D activity from cabinet- and agency-level budgets down to the program, project, and award levels. CTI made the beta version of the RaDiUS database to the Office of Science and Technology Policy and other parts of the federal government. When completed, RaDiUS will be accessible to designated users via the Internet through the World Wide Web.
federal R&D are likely to remain strong. We believe that this template can help guide CENR through the process of developing effective performance measures for their programs, and provide a menu of options for choosing specific measures. Perhaps assisted by new information technologies such as RaDiUS, CENR should be able to develop a evaluation process consistent with a creative and valuable R&D enterprise and the demands for increased accountability in the investment of the publics' funds.
BIBLIOGRAPHY


### TABLE 1: A HIERARCHY OF GOALS

<table>
<thead>
<tr>
<th>STRATEGIC PLAN</th>
<th>IMPLEMENTATION PLAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) [National goals]</td>
<td>4) Program goals and plans</td>
</tr>
<tr>
<td>2) [Environmental policy] goals</td>
<td>Agency roles and accomplishments</td>
</tr>
<tr>
<td>Policy questions and issues</td>
<td>Programmatic milestones/schedule</td>
</tr>
<tr>
<td>3) Scientific [Research and Development] goals</td>
<td>Constraints</td>
</tr>
<tr>
<td>Current state of understanding</td>
<td>Performance measures, metrics, and evaluation criteria</td>
</tr>
<tr>
<td>Research priorities</td>
<td>Resources</td>
</tr>
<tr>
<td>External input</td>
<td>External input</td>
</tr>
<tr>
<td>International dimension</td>
<td>International dimension</td>
</tr>
</tbody>
</table>

Note: Elements of this table, except those in [], are from the CENR Subcommittees' Strategy and Implementation Plans Guidance, July 11, 1994.
TABLE 2: Performance Measures for Individual Programs

<table>
<thead>
<tr>
<th>PERFORMANCE TOWARDS GOALS</th>
</tr>
</thead>
</table>

**Ex-Ante Evaluation**
- Peer Review of relevance (NSF) [S/D/T/A]
- Programs with user approval (IRI-10) [D/T/A]
- Percentage outside funding (IRI-12) [T]

**Ex-Post Evaluation**
- Program evaluation (NSF) [S/D/T/A]
- Retrospective studies (NSF) [S/D/T/A]
- Accomplishment books (NSF) [S/D/T/A]
- Users' ratings of benefits (IRI-15) [S/D/T/A]
- Social return on investment (IRI-1) [D/T/A]
- Impacts of research (IRI-15) [D/T/A]
- Timeliness of research (IRI-16 & 19) [A]

**PROCESS EVALUATION**

<table>
<thead>
<tr>
<th>Process Evaluation</th>
</tr>
</thead>
</table>

- Peer review of quality (NSF/IRI-23) [S/D/T/A]
- Quality of personnel (IRI-18) [S/D/T/A]
- Goal Clarity (IRI-30) [S/D/T/A]
- Morale (IRI-29) [S/D/T/A]
- Use of Milestones (IRI-11) [D/T/A]
- Cross Functional Teams (IRI-14) [D/T/A]

- Technical reviews (NSF/IRI-20) [S/D/T/A]
- Publication counts (NSF/IRI-27) [S]
- Citation counts (NSF) [S]
- Milestones achieved (IRI-25) [D/T/A]
- Contact time with users (IRI-26) [T/A]
- Technology transfer (IRI-13) [T]
- Patents (IRI-21) [T]
### TABLE 3: Usefulness of IRI Metrics to Stakeholders

<table>
<thead>
<tr>
<th></th>
<th>Value Creation</th>
<th>Integration w Business</th>
<th>Portfolio M'gt</th>
<th>Practice of R&amp;D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Board members &amp; Financial Community</td>
<td>3</td>
<td>1.1</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>CEO</td>
<td>3</td>
<td>2.4</td>
<td>2.0</td>
<td>1.4</td>
</tr>
<tr>
<td>Business Management</td>
<td>2.6</td>
<td>3</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>R&amp;D Management</td>
<td>2</td>
<td>2.9</td>
<td>3.0</td>
<td>2.9</td>
</tr>
<tr>
<td>R&amp;D Staff</td>
<td>0.7</td>
<td>1.3</td>
<td>1.0</td>
<td>2.5</td>
</tr>
</tbody>
</table>

3 = maximum utility; 0 = no utility

### TABLE 4: Performance Measures for Program Balance

<table>
<thead>
<tr>
<th>PERFORMANCE TOWARDS GOALS</th>
<th>Ex-Ante Evaluation</th>
<th>Ex-Post Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Projected value of research (IRI-2) [S/D/T/A]</td>
<td>• Financial returns (IRI-1) [S/D/T/A]</td>
</tr>
<tr>
<td></td>
<td>• Strategic Alignment (IRI-7) [T]</td>
<td>• Users' rating of benefits (IRI-15) [S/D/T/A]</td>
</tr>
<tr>
<td></td>
<td>• Distribution of Investments (IRI-8) [T]</td>
<td>• Ways research exploited (IRI-9) [D/T/A]</td>
</tr>
</tbody>
</table>
National Science & Technology Council (NSTC)

Executive Secretariat

NSTC Deputies Group

Committee on Health, Safety, & Food R&D
Committee on Information & Communication R&D
Committee on National Security
Committee on Civilian Industrial Technology

Committee on Fundamental Science
Committee on International Science Engineering & Technology
Committee on Environment & Natural Resources
Committee on Transportation R&D
Committee on Education & Training R&D

National Science and Technology Council Committee on Environment and Natural Resources Executive Committee

Environmental Technology
Risk Assessment
Social & Economic Science
Observations and Data Management Task Group
Ecosystems Work Group

Air Quality
Biodiversity & Ecosystems
Global Change
Natural Disaster Reduction

Resource Use & Management
Toxic Substances/Hazardous & Solid Waste
Water Resources/Coastal & Marine Environments

Figure 1
LEVEL 2: ENVIRONMENTAL POLICY GOAL
Protect the environment for present and future generations while enabling development to proceed in a sustainable manner

POLICY QUESTION: How will the environment behave in response to various actions?

LEVEL 3: R&D GOAL
Develop scientifically sound, user friendly information and validated analytical models to enable exploration of possible consequences of alternative management and policy decisions

RESEARCH QUESTION: What is the current situation?

LEVEL 4: SCIENCE PROGRAM GOAL:
Understand current situation (snapshot and trends)

discover, describe, inventory, and classify the world's organisms to provide the fundamental information base for monitoring and surveying biological diversity

LEVEL 4: SCIENCE PROGRAM GOAL:
Develop Models

(a) understand the environmental, ecological, and evolutionary processes that generate and maintain biological diversity, sustain viable populations of species, and support ecosystem structure and function

(b) determine status, change, and trends in biodiversity and ecosystem function

LEVEL 4: ASSESSMENT PROGRAM GOAL:
Assess and improve models to manage ecosystems

obtain (and improve) the scientific information base required for sound ecosystem management

1.1 Classification and Systematic
1.1.1 Taxonomic Authority System
1.1.2 Systematic Training Initiative
1.1.3 Museum Support Initiative

1.2 Inventory and Monitoring
1.2.1 Ecological Indicators
1.2.1 Biological Catalogs
1.2.3 Ecosystem Mapping
1.2.4 Species Inventories

2.1 National Network of Ecosystem Sites
2.1.1 Ecosystem Classification
2.1.2 Network of Ecosystem Sites
2.2 Long-term and Large-scale Characterization of Impacts on Biodiversity and Ecosystem Patterns
2.2.1 Long-term Trends and Patterns

etc. @ @ @

4.1 Adaptive Management
4.1.1 Adaptive Management Demonstrations
4.1.2 Technology Transfer to Private Users
APPENDIX A:
AUGUST 1994 VERSION OF THE TEMPLATE
Performance Measures for the Federal Investment in Science and Technology:

A Draft Template for CENR

August 9, 1994

The Critical Technologies Institute (CTI) is assisting the Office of Science and Technology Policy (OSTP) to develop performance goals and milestones for the Federal investments in science and technology. As part of this effort, CTI is working with the Committee on Environment and Natural Resources Research (CENR) and two of its sub-committees to develop a template for performance measures for environment and natural research research. This template will assist the CENR sub-committees to develop the performance measures for their strategic and implementation plans, as called for in the CENR guidance. The final template will also assist CENR and OSTP to integrate the strategic and implementation plans from all of the CENR subcommittees, as well as explain to the larger community how CENR used performance measures in these plans. This current draft is an initial version of the CTI template for consideration and comments by members of OSTP and CENR.
What Should Performance Measures Accomplish?

Performance measures should:
- Help OMB and Congress determine
  - The goals of the federal investment in S&T
  - Where the money goes
  - What society gets from its investment

Encourage a balanced CENR program:
- scientifically excellent
- policy-relevant
- meeting near- and long-term policy and science needs

Performance measures help determine the efficiency and effectiveness of federal R&D programs. These measures include milestones, quantitative metrics, and evaluation techniques of all kinds. In this briefing we will suggest the types of performance measures most appropriate for different programs of environment and natural resources research.

It is first important to specify what these performance measures are intended to accomplish, since the choice of performance measures will be strongly shaped by their intended purposes. CENR was established to ensure that federal environmental R&D policies and programs contribute effectively to national goals and to increase the overall effectiveness and productivity of these programs. Performance measures can help evaluate the contributions R&D programs make towards the ultimate goals of their sponsors.

There is currently much interest in measuring the returns to society from federal programs of all kinds -- not just R&D -- as evidenced by the National Performance Review and the Government Performance and Results Act of 1993. Because the returns from R&D are usually much harder to evaluate than those from other federal activities, CENR is also charged with ensuring a balanced and comprehensive environmental R&D program. Performance measures can help CENR assure its research satisfies the criteria set forth by Robert Watson and Jim Baker in their May 4 testimony to the House Committee on Science, Space, and Technology -- that it be scientifically excellent; policy-relevant, not policy-driven; and meet the near- and long-term needs of the policy and scientific communities.
The sub-committees of the CENR have been given guidance on writing strategic and implementation plans for the research under their purview. CTI has focused on a four-level hierarchy contained in the guidance, which links broad national goals (as stated in the Gibbons-Fanetta memo on the federal R&D budgets for FY 96), environmental policy goals, science goals, and program goals. This hierarchy is useful because performance measures are best applied to levels of the hierarchy, or to the links between different levels.

This slide shows the hierarchy within the CENR guidance. The boxes indicate the four levels of the hierarchy and the oval indicates the performance measures which are the subject of this briefing. Performance measures is the general term which includes programmatic milestones, science and policy-relevant milestones, and evaluation criteria of all types.

Note three elaborations on the CENR guidance. We have explicitly included broad national goals as the first level of the hierarchy although these goals will likely appear only implicitly in the subcommittees’ strategic plans. We also focus our attention on environmental policy goals within level 2 and refer to scientific and other R&D goals in level 3, because we will differentiate among different types of research, and the appropriate performance measures for each, later in this briefing.
Performance measures are appropriate for the program goals (level 4) in the hierarchy and to the links between all the levels. This slide is the first of three which shows how this can be done. Here we lay out the levels of the hierarchy identified on the previous slide, from national goals, environmental policy goals, R&D goals, to program goals. Environmental policy goals answer the question "what actions do the federal government and other actors need to take to move us towards our national goals?" R&D goals answer the question, "what information does the federal government and these other actors need to carry out the necessary actions successfully?" The Policy Questions identified in the guidance are the bridge between these goals. Program goals answer the question "how do we get this information?" Program goals may be thought of as questions which could be answered by a series of RFPs.

We have divided program goals into four categories -- science, integrated assessments, technology, and data collection and management programs. As we will explain in detail below, there is a different set of performance measures most appropriate for each of these categories. Not all CENR subcommittees will necessarily have research in each of these categories. Some subcommittees may also have additional types of research.
**Definitions**

Performance Measures - methods for measuring the efficiency or effectiveness of programs, including metrics, evaluation techniques, and milestones.

Program Goal - an important set of R&D activities identified in a CENR implementation plan to be carried out one or more agencies.

Performance Towards Goals - performance measures which evaluate the extent to which a program contributes to its goals or to higher-level goals.

Process Evaluation - performance measures which evaluate the efficiency of a program (its productivity and managerial performance).

Programmatic Milestones - target for completion of significant activity in (R&D) program; includes a timetable component.

There are many different terms with many different meanings used by those developing performance measures for R&D programs. In this slide we specify the definitions of some of the key terms used in this template. They are drawn from the literature on R&D evaluation, and definitions used in the Government Performance and Results Act and recent appropriations bills.

As we will discuss, **performance measures** is the general term for methods of measuring the efficiency or effectiveness of the activities which make up the **program goals** in the CENR implementation plans. Performance measures can also evaluate the balance among program goals.

**Performance towards goals** and **process evaluations** are two types of performance measures, which will be described in the next slide.

**Programmatic milestones** are one type of process evaluation.
Assessing ‘Performance Towards Goals’ Preferable, But Not Always Possible

<table>
<thead>
<tr>
<th>PERFORMANCE MEASURE</th>
<th>AUDIENCE</th>
<th>APPLICABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERFORMANCE TOWARDS GOALS</td>
<td>Policy-Makers (e.g., OMB, Congress) and Program Managers</td>
<td>Difficult for programs (esp. basic sciences) whose desired outcomes can not be clearly defined</td>
</tr>
<tr>
<td>Assessing how well the program contributes to higher level goals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROCESS EVALUATION</td>
<td>Primarily Program Managers</td>
<td>All programs</td>
</tr>
<tr>
<td>Assessing how well the program is being carried out</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some performance measures directly evaluate the extent to which a program contributes to its objectives or to higher-level goals. We label these “performance toward goals.” Such measures are preferable when available, since they are immediately useful to policy-makers such as OMB and Congress, who must best allocate the resources available to achieve broad national goals.

However, it is often difficult to assess the contributions of R&D programs because no one can predict what the outcomes of the programs may be. We do not know before hand when and what knowledge a particular program will produce, and in some cases it may take years after the program is completed to properly assess its significance.

When “performance toward goals” measures are unavailable, we can use measures of program efficiency, which we label “process evaluation”. Programmatic milestones are an important example of process evaluations, as are benchmarking and counting outputs such as scientific papers. Process evaluations are directly useful to managers trying to run cost-effective and timely programs. Their usefulness for policy-makers is more indirect, requiring the assumption that a well-run program is likely making progress towards the ultimate objectives.
This slide indicates that the subcommittees can use both process evaluation and performance towards goals measures to evaluate individual programs in their implementation plans. As shown here, most programs can be evaluated for their efficiency (process evaluation) and for their contribution to level 3 and to level 2 goals (performance towards goals). At present, science programs rely far more heavily on process evaluation only.
Government and industry have gathered much experience with performance measures for individual R&D programs. There is also a significant academic literature on such measures. Drawing on all these sources, this slide lists many of the performance measures CENR subcommittees can use in their implementation plans for evaluating four different types of R&D — science, integrated assessments, technology, and data collection and management.

We distinguish between ex-ante performance measures (evaluation before the completion of, and during, a program) and ex-post performance measures (evaluation after completion of a program). Ex-ante measures are desirable, but usually less reliable. We also distinguish between 'process evaluations' (italics) and 'performance towards goal's measures (plain text). Note that all R&D categories except science have available relatively good ex-post measures of their contributions to higher level goals, and that milestones and benchmarking provide useful ex-ante measures. Process evaluations dominate the available measures for science.

This chart will help CENR sub-committees choose appropriate performance measures for different programs within their implementation plans. It suggests where programmatic, scientific, and policy milestones are possible, and the types of evaluation tools which can be used. CTI is currently preparing detailed reviews of each of these measures. In the interim, subcommittees are invited to contact CTI staff members to discuss the use of these performance measures in their implementation plans.
A set of performance measures for individual programs will in general be insufficient to achieve all the purposes CENR intends. In particular, our ability to construct meaningful 'performance towards goals' measures is currently too primitive to ensure a balanced program of CENR research. Thus we will also need performance measures which directly evaluate the balance among programs. These measures will help ensure that the portfolio of level 4 programs can contribute effectively to achieving our level 3 R&D goals, and that the portfolio of level 3 R&D goals can contribute effectively to achieving our level 2 environmental policy goals. For instance, these measures could help maintain an appropriate balance between near-term and far-term research.
### Possible Performance Measures for Program Balance

<table>
<thead>
<tr>
<th>Portfolio Analysis</th>
<th>Ex-Ante Evaluation</th>
<th>Ex-Post Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does the portfolio contain a robust mix of programs?</td>
<td>- Expert judgments</td>
<td>- Use by policy-makers</td>
</tr>
<tr>
<td></td>
<td>- Gap-overlap analysis</td>
<td>- Stakeholder assessments</td>
</tr>
<tr>
<td></td>
<td>- Role of the blind</td>
<td>- Retrospective assessment of policy impacts, &quot;train wrecks avoided.&quot;</td>
</tr>
<tr>
<td></td>
<td>- Benchmarking</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value of Information</th>
<th>Ex-Ante Evaluation</th>
<th>Ex-Post Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>What information, if available and used, will make the most difference to policy outcome? When is the information needed?</td>
<td>- Expert judgments</td>
<td>- Use by policy-makers</td>
</tr>
<tr>
<td></td>
<td>- Stakeholder inputs</td>
<td>- Stakeholder assessments</td>
</tr>
<tr>
<td></td>
<td>- Multiattribute survey techniques</td>
<td>- Retrospective assessment of policy impacts, &quot;train wrecks avoided.&quot;</td>
</tr>
<tr>
<td></td>
<td>- Policy modeling</td>
<td></td>
</tr>
</tbody>
</table>

There is much less experience with performance measures for the balance among program balance than there is with measures for individual R&D programs. The methods generally fall into two broad groups. The first, which we label 'portfolio analysis,' evaluates whether the portfolio of programs or goals appears well-balanced in terms of broad categories, for example near- vs. far-term activities, high-risk/high-payoff vs. low-risk/low-payoff activities, and physical vs. social science research. The second group, which we label 'value of information,' attempts to determine the mix and timing of information which, if available and acted upon, would have the most beneficial impact on the actions of policy-makers.

Most of the measures for ex-ante 'portfolio analysis' evaluations tend to depend on implicit or explicit correlations of ex-post to ex-ante performance for prior similar programs. That is, the measures require an assumption that, if one compares the current programs to prior programs, and finds similarity along the measured dimensions, then the current programs are likely to be successful. 'Value of information' measures use a formal methodology to focus on the particular goals in question, but tend to be very difficult to apply with fidelity.

As with the performance measures for individual programs, CTI is currently preparing more detailed information on these measures of program balance. In the interim, subcommittees are invited to contact CTI staff members to discuss the use of these measures in their implementation plans.
This slide reviews how the CENR subcommittees can use the information presented here to help them prepare their implementation plans.

First, the subcommittees should review the environmental policy goals (level 2) and the R&D goals (level 3) which support them, as laid out in their strategic plans. The arguments justifying their research priorities, called for by the strategic plan guidance, will likely draw upon arguments similar to the performance measures for program balance discussed on Slide #10.

Next, the subcommittees should review the program goals (level 4) they will include in their implementation plans. They should divide these goals in categories, such as the science, integrated assessment, technology, data collection and management categories we use here, or other groupings as the subcommittee finds appropriate.

The subcommittees then can choose performance measures for the individual programs from the list on Slide #8. Ex-ante ‘performance towards goals’ measures should be used where possible, with other types used as necessary. In practice, each program will likely use a mix of measures of different types.

Finally, the subcommittees can chose measures for program balance, from the list on Slide #10, to help justify the choice and mix of different types of program goals (level 4), and to help evaluate the contribution of these programs to R&D (level 3) and environmental policy (level 2) goals.
APPENDIX B:
INDUSTRIAL RESEARCH INSTITUTE (IRI) METRICS
INDUSTRIAL RESEARCH INSTITUTE
WORKSHOP

MONDAY AFTERNOON
10/24/94

TECHNOLOGY VALUE PYRAMID℠

MENU OF METRICS
AND
DEFINITIONS FOR THE MENU
# The Menu of Metrics

## R&D Metrics and Their Assignments to Particular Levels in the Technology Value Pyramid

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value Creation</th>
<th>Portfolio Assessment</th>
<th>Integration with the Business</th>
<th>Value of the Technical Assets</th>
<th>Practice of R&amp;D</th>
</tr>
</thead>
</table>
| 1. Financial Return  
a. New Sales Ratio  
b. Cost Savings Ratio  
c. R&D Yield  
d. R&D Return | x | x | | | |
| 2. Projected Value of the R&D Pipeline  
a. Projected Sales Value from Pipeline  
b. Projected Income Value from Pipeline | x | x | | | |
| 3. Comparative Manufacturing Cost | x | | | | x |
| 4. Product Quality & Reliability  
a. Customer/Consumer Evaluation  
b. Reliability/Defect Rate Assessment | x | | x | | |
| 5. Gross Profit Margin | | x | | | |
| 6. Market Share  
a. Direct Market Share  
b. Related Market Share | x | | | | |
| 7. Strategic Alignment  
a. Corporate & Business Unit  
b. Goal Coverage | x | x | | x | x |
# The Menu of Metrics

R&D metrics and their assignments to particular levels in the Technology Value Pyramid

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value Creation</th>
<th>Portfolio Assessment</th>
<th>Integration with the Business</th>
<th>Value of the Technical Assets</th>
<th>Practice of R&amp;D</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Distribution of Technology Investment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. New to World</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Established</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Maintaining position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Base</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. Key</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Pacing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. External Technology Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h. Core Technical Competencies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>9. Number of Ways Technology is Exploited (7)</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Number of Project Definitions Having Business/Marketing Approval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Use of Project Milestone System</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>12. Percent Funding by the Business</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Technology Transfer to Manufacturing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Use of Cross-Functional Teams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>15. Rating of Product Technology Benefits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Customer Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Economic Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Market Share Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Response Time to Competition's Moves</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# The Menu of Metrics

## R&D Metrics and Their Assignments to Particular Levels in the Technology Value Pyramid

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value Creation</th>
<th>Portfolio Assessment</th>
<th>Integration with the Business</th>
<th>Value of the Technical Assets</th>
<th>Practice of R&amp;D</th>
</tr>
</thead>
<tbody>
<tr>
<td>17. Current Investment in Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 18. Quality of Personnel  
  a. Internal Customer Rating  
  b. External Customer Rating  
  c. External Recognition  
  d. Published Works |               |                      |                               | x                |                 |
| 19. Development Cycle Time  
  a. Market Cycle Time  
  b. Project Management Cycle Time |               |                      | x                             | x                | x               |
| 20. Customer Rating of Technical Capability |               |                      |                               |                               |                 |
| 21. Number & Quality of Patents  
  a. Percent Useful  
  b. Value Ratio  
  c. Retention Percent  
  d. Cost of Invention |               |                      | x                             | x                | x               |
| 22. Sales Protected by Proprietary Position  
  a. Percent Patent Protected Sales  
  b. Percent Proprietary Sales |               |                      | x                             | x                |                 |
| 23. Peer Evaluation  
  a. External  
  b. Internal |               |                      |                               | x                | x               |
## The Menu of Metrics

**R&D Metrics and their Assignments to Particular Levels in the Technology Value Pyramid**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value Creation</th>
<th>Portfolio Assessment</th>
<th>Integration with the Business</th>
<th>Value of the Technical Assets</th>
<th>Practice of R&amp;D</th>
</tr>
</thead>
<tbody>
<tr>
<td>24. Customer Satisfaction</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>a. External</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>b. Internal</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>25. Development Pipeline Milestone Achieved</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>a. Percent of Project Milestones Achieved</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>b. Performance Level at Each Milestone</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>26. Customer Contact Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>27. Preservation of Technical Output</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>28. Efficiency of Internal Technical Process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>a. Project Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Portfolio Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. Employee Morale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>30. Goal Clarity</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>31. Project Ownership/Empowerment</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>32. Management Support</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>33. Project Championship</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>
## DEFINITIONS FOR THE MENU

<table>
<thead>
<tr>
<th>METRIC</th>
<th>DEFINITION</th>
<th>CATEGORY</th>
</tr>
</thead>
</table>
| **1. FINANCIAL RETURN**       | **a. New Sales Ratio (NSR)**: The ratio of sales revenue in year "i" from product developments commercialized in years "i-j" to "i-1" to total sales revenue in year "i", or:  
                              | \[ NSR = (1/Sales) \cdot \sum_{N=i-j}^{N=i-1} (NP \text{ Sales}) \]  
                              | where: \( NP \text{ Sales} = \text{sales revenue in year "i" from product developments commercialized in year N} \) \( \text{Sales} = \text{total sales revenue in year "i"} \) | VC       |
|                               | **b. Cost Savings Ratio (CSR)**: The ratio of savings in the cost of goods sold in year "i" from process developments or product changes adopted in years "i-k" to "i to 1" to average gross profits in year "i", or:  
                              | \[ CSR = (1/GP) \cdot \sum_{N=i-k}^{N=i-1} \text{CG Savings} \]  
<pre><code>                          | where: \( \text{CG Savings} = \text{reduction in cost of goods sold in year &quot;i&quot; from process developments or product changes adopted in year N} \) \( \text{GP} = \text{average gross profit for business unit in year &quot;i&quot;} \) | VC       |
</code></pre>
<table>
<thead>
<tr>
<th>METRIC</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td><strong>R&amp;D Yield</strong></td>
</tr>
<tr>
<td></td>
<td>The gross profit (GP) contribution from the sale of new and improved products (NP) and from the lower cost of goods from new and improved processes or new formulations (CR).</td>
</tr>
<tr>
<td></td>
<td><strong>where:</strong></td>
</tr>
<tr>
<td></td>
<td>NP = GP * NSR  \quad CR = GP * CSR</td>
</tr>
<tr>
<td></td>
<td><strong>or:</strong></td>
</tr>
<tr>
<td></td>
<td>R&amp;D Yield = GP (NSR + CSR)</td>
</tr>
<tr>
<td>6.</td>
<td><strong>R&amp;D Return</strong></td>
</tr>
<tr>
<td></td>
<td>The ratio of R&amp;D benefits to R&amp;D investment.</td>
</tr>
<tr>
<td></td>
<td><strong>R&amp;D Return = R&amp;D Yield/R&amp;D Effort</strong></td>
</tr>
<tr>
<td></td>
<td><strong>R&amp;D Effort = Annual expenditure on R&amp;D.</strong></td>
</tr>
<tr>
<td></td>
<td><strong>where:</strong></td>
</tr>
<tr>
<td></td>
<td>R = R&amp;D Effort</td>
</tr>
<tr>
<td></td>
<td>I = R&amp;D Intensity</td>
</tr>
<tr>
<td></td>
<td>Sales</td>
</tr>
<tr>
<td></td>
<td>R = R&amp;D Effort/Sales</td>
</tr>
</tbody>
</table>
### DEFINITIONS FOR THE MENU

<table>
<thead>
<tr>
<th>METRIC</th>
<th>DEFINITION</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. PROJECTED VALUE OF THE R&amp;D PIPELINE</td>
<td>a. Projected Sales Value from Pipeline: Fraction of future sales by year projected from projects in the R&amp;D pipeline, incorporating probability of attaining objective for each project.</td>
<td>VC; PA</td>
</tr>
<tr>
<td></td>
<td>b. Projected Income Value from Pipeline: Fraction of future net income (and/or return) by year projected from projects in the R&amp;D pipeline, incorporating NPV times probability of attaining objective for each project.</td>
<td></td>
</tr>
<tr>
<td>3. COMPARATIVE MANUFACTURING COST</td>
<td>Benchmarked manufacturing cost data vs. competition for substantially identical manufacturing steps or for producing substantially the same product (eliminating factors unrelated to technology-based differences).</td>
<td>VC; VTA</td>
</tr>
<tr>
<td>4. PRODUCT QUALITY &amp; RELIABILITY</td>
<td>a. Customer or Consumer Evaluation: Relative quality (as evaluated by the customer/consumer) vs. competitive product in blinded product evaluation by techniques appropriate to the industry segment.</td>
<td>VC; VTA</td>
</tr>
<tr>
<td></td>
<td>b. Reliability/Defect Rate Assessment: At the firm level, fraction of company's product output that meets or exceeds the established quality standards. At the product level, fraction of a given product's quarterly output that meets or exceeds the established quality standards.</td>
<td></td>
</tr>
<tr>
<td>5. GROSS PROFIT MARGIN</td>
<td>Gross Profit as a percentage of sales, where gross profit equals net sales minus cost of goods sold (product costs plus direct manufacturing costs). Value assessment should be based on change in gross profit margin from period to period. (Periods should be appropriate to an industry and may be in excess of 1 year.)</td>
<td>VC; VTA</td>
</tr>
</tbody>
</table>
## DEFINITIONS FOR THE MENU

<table>
<thead>
<tr>
<th>METRIC</th>
<th>DEFINITION</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. <strong>MARKET SHARE</strong></td>
<td><strong>a. Direct Market Share</strong> Company (or business unit) market share in various categories measured as appropriate for the industry or category. As with gross profit margin, changes should be assessed at least annually to determine rate of progress or decline.</td>
<td>VC; VTA</td>
</tr>
<tr>
<td></td>
<td><strong>b. Related Market Share</strong> Share data in related markets for indication of threats or opportunities.</td>
<td></td>
</tr>
<tr>
<td>7. <strong>STRATEGIC ALIGNMENT</strong></td>
<td><strong>a. Corporate and Business Unit</strong> Fraction of total R&amp;D portfolio which is consistent with corporate goals, and, where applicable, for each major business unit. Should also include specific identification of the corporate or business goal with which a project is identified.</td>
<td>PA; IW; VTA</td>
</tr>
<tr>
<td></td>
<td><strong>b. Goal Coverage</strong> Fraction of corporate or business unit goals requiring technology development that are addressed by the R&amp;D portfolio.</td>
<td></td>
</tr>
<tr>
<td>8. <strong>DISTRIBUTION OF TECHNOLOGY INVESTMENT</strong></td>
<td>Analysis of the fraction of the total R&amp;D investment along various dimensions. Each company should describe its portfolio of technical activities along dimensions which are important to decision making and communication between stakeholders. Some of the more common dimensions are:</td>
<td>PA; VTA</td>
</tr>
<tr>
<td></td>
<td><strong>A. Dimensions potentially of greater interest to the CEO (and to Business management):</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Categorization of Reward vs. Risk</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- By Product Line or Business Unit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- For Maintenance of current business, Expansion of current business, or Creation of New Business</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Environmentally Driven vs. non-Environmental</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Distribution according to Time of Commercialization</td>
<td></td>
</tr>
</tbody>
</table>
## DEFINITIONS FOR THE MENU

<table>
<thead>
<tr>
<th>METRIC</th>
<th>DEFINITION</th>
<th>CATEGORY</th>
</tr>
</thead>
</table>
| B. Dimensions potentially of greater interest to Business Management (and some CEOs) | - For cost reduction, applications development, or performance differentiation  
- For Current markets, Markets new to the company, or Markets New to the World  
- For Current Technology, Technology new to the Company, or Technology New to the World |  |
| C. Dimensions potentially of greater interest to R&D Management | - Distribution by project stage  
- Distribution by technical discipline  
- External R&D vs. Internal  
- Base, Key or Pacing Technology  
- Core competencies vs. new competencies  
- U.S. vs. non-U.S. |  |

9. **NUMBER OF WAYS TECHNOLOGY IS EXPLOITED**  
Assessment of the number of different product types or business segments utilizing or planning to utilize a given technical asset (the exploitation of functional competencies).

10. **NUMBER OF PROJECT DEFINITIONS HAVING BUSINESS/MARKETING APPROVAL**  
Fraction (or percent) of projects in the total R&D portfolio with explicit business unit and/or corporate business management sign-off.
## Definitions for the Menu

<table>
<thead>
<tr>
<th>Metric</th>
<th>Definition</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. <strong>USE OF PROJECT MILESTONE SYSTEM</strong></td>
<td>Fraction of projects in the total portfolio going through a defined project management system with defined milestones.</td>
<td>PRD; IWB</td>
</tr>
<tr>
<td>12. <strong>PERCENT FUNDING BY THE BUSINESS</strong></td>
<td>Fraction (or percent) of the R&amp;D budget from business unit sources.</td>
<td>IWB</td>
</tr>
<tr>
<td>13. <strong>TECHNOLOGY TRANSFER TO MANUFACTURING</strong></td>
<td>Semi-quantitative assessment (e.g., an interval scale from 1 to 5) of the effectiveness of the transfer with ratings obtained from both R&amp;D and Manufacturing sides of the interface.</td>
<td>IWB; PRD</td>
</tr>
<tr>
<td>14. <strong>USE OF CROSS-FUNCTIONAL TEAMS</strong></td>
<td>Fraction of projects in the R&amp;D portfolio with specific cross-functional teams assigned. This analysis can be further subdivided by project types, e.g., short-term development, long-term development, process development, applied research, etc.</td>
<td>IWB; PRD</td>
</tr>
<tr>
<td>15. <strong>RATING OF PRODUCT TECHNOLOGY BENEFITS</strong></td>
<td>Numerical ranking of a firm's product by a given customer divided by that customer's ranking of the best competitive product. This ratio can be averaged across customers and/or market segments using the product to obtain an average ratio value.</td>
<td>VTA; VC</td>
</tr>
<tr>
<td>a. <strong>Customer Rating</strong></td>
<td>(Price differential per unit obtained by virtue of quality feature(s) derived from technical effort minus the cost per unit of providing the feature(s)) times unit sales volume for products containing the feature(s).</td>
<td></td>
</tr>
<tr>
<td>b. <strong>Economic Evaluation</strong></td>
<td>Differential share gain(s) at a constant price for product(s) containing the quality feature(s).</td>
<td></td>
</tr>
</tbody>
</table>
# DEFINITIONS FOR THE MENU

<table>
<thead>
<tr>
<th>METRIC</th>
<th>DEFINITION</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. RESPONSE TIME TO COMPETITIVE</td>
<td>Time required for the firm to match competitors’ newest product benefit(s) divided by time required for competitor to match the firm’s newest product benefit(s).</td>
<td>VTA; PRD</td>
</tr>
<tr>
<td>MOVES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. CURRENT INVESTMENT IN TECHNOLOGY</td>
<td>Current annual expenditures for R&amp;D staff and equipment ratioed to best competitor, to industry average, and to industry total.</td>
<td>VTA</td>
</tr>
<tr>
<td>18. QUALITY OF PERSONNEL</td>
<td>Several measures are possible; the measure or measures chosen should be appropriate for the firm’s technology development strategy.</td>
<td>VTA, PRD</td>
</tr>
<tr>
<td>a. Internal Customer Rating</td>
<td>Interval rating scale (e.g., a scale from 1 to 5) from R&amp;D’s internal customers, such as Marketing, Manufacturing, etc.</td>
<td></td>
</tr>
<tr>
<td>b. External Customer Rating</td>
<td>Interval rating scale from the firm’s major customers.</td>
<td></td>
</tr>
<tr>
<td>c. External Recognition</td>
<td>External awards and invited lectures received by the professional staff over a relevant time period.</td>
<td></td>
</tr>
<tr>
<td>d. Published Works</td>
<td>Publications and patents by the professional staff over a relevant time period.</td>
<td></td>
</tr>
<tr>
<td>19. DEVELOPMENT CYCLE TIME</td>
<td>Elapsed time from identification of a customer product need until commercial sales commence.</td>
<td>VTA; PRD</td>
</tr>
<tr>
<td>a. Market Cycle Time</td>
<td>Elapsed time from establishment of a discrete project to address an identified customer product need until commercial sales commence.</td>
<td></td>
</tr>
<tr>
<td>b. Project Management Cycle Time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For both (a) and (b) the end point can be time when manufacturing feasibility is established for those cases where no commercialization occurs. Compare to historical values and benchmark vs. competition, if possible. Group by categories of projects (e.g., major new product, minor product variation, etc.) Can also be used to track milestone attainment rate for firms using a stage gate management process.
## DEFINITIONS FOR THE MENU

<table>
<thead>
<tr>
<th>METRIC</th>
<th>DEFINITION</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. CUSTOMER RATING OF TECHNICAL CAPABILITY</td>
<td>Average customer rating (internal or external) of overall technical capability of the firm (interval rating scale) in providing technical service and/or new product innovations. Can be ratioed to ratings for relevant competitors for benchmarking purposes.</td>
<td>VTA</td>
</tr>
<tr>
<td>21. NUMBER AND QUALITY OF PATENTS</td>
<td></td>
<td>VTA; PRD</td>
</tr>
<tr>
<td>a. Percent Useful</td>
<td>Percentage of active patents from the company's total patent estate which are incorporated into or used to defend the firm's commercial products or processes.</td>
<td>VTA</td>
</tr>
<tr>
<td>b. Value Ratio</td>
<td>Interval rating (1 to 5) for potential strategic value times rating (1 to 5) for strength of protection divided by 25 (maximum attainable value). Yields a number between 0 and 1.</td>
<td>VTA</td>
</tr>
<tr>
<td>c. Retention Percent</td>
<td>Percent of granted patents maintained.</td>
<td>VTA</td>
</tr>
<tr>
<td>d. Cost of Invention</td>
<td>Number of patents from R&amp;D/R&amp;D effort. One can also calculate this just using the number of useful patents from R&amp;D.</td>
<td>VTA</td>
</tr>
<tr>
<td>22. SALES PROTECTED BY PROPRIETARY POSTION</td>
<td></td>
<td>VTA</td>
</tr>
<tr>
<td>a. % Patent Protected Sales</td>
<td>Percentage of sales protected by patents owned by the company.</td>
<td>VTA</td>
</tr>
<tr>
<td>b. % Proprietary Sales</td>
<td>Percentage of sales protected by patents plus trade secrets and/or other exclusive company know-how or arrangements.</td>
<td>VTA</td>
</tr>
<tr>
<td>23. PEER EVALUATION</td>
<td></td>
<td>VTA; PRD</td>
</tr>
<tr>
<td>a. External</td>
<td>Numerical rating (interval rating scale) from 1 to 5 by a panel of external experts on the merits of the firm's technology positioning and technology management practices. Panel selections critical; it must be capable and objective. Could include outside directors from technology companies or university science/engineering departments, consultants in technology management &amp; strategy, appropriate university professors, venture capitalists, etc. Same rating scale applied by internal experts; probably staff on company's technical ladder.</td>
<td>VTA</td>
</tr>
<tr>
<td>b. Internal</td>
<td></td>
<td>VTA</td>
</tr>
</tbody>
</table>

viii
## DEFINITIONS FOR THE MENU

<table>
<thead>
<tr>
<th>METRIC</th>
<th>DEFINITION</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>24. CUSTOMER SATISFACTION</td>
<td></td>
<td>VTA; PRD</td>
</tr>
<tr>
<td>a. External</td>
<td>Average rating by key external customers using a 1 to 5 interval rating scale to evaluate various dimensions regarding product technology or process technology benefits and technical service provided.</td>
<td></td>
</tr>
<tr>
<td>b. Internal</td>
<td>Same rating approach along dimensions important to key internal customers such as Marketing, Engineering, Manufacturing etc.; dimensions would include, for example, timelines of developments, competitiveness of solutions developed, etc.</td>
<td></td>
</tr>
<tr>
<td>25. DEVELOPMENT PIPELINE</td>
<td></td>
<td>PRD</td>
</tr>
<tr>
<td>MILESTONES ACHIEVED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Percent of project milestones achieved</td>
<td>Percent of project milestones achieved within three months of projected achievement date (or within that time appropriate for an industry). Plot as histogram to reveal actual performance, showing percent that beat the target as well as those that miss it.</td>
<td></td>
</tr>
<tr>
<td>b. Performance Level at each milestone</td>
<td>Percent completion of objectives expected by a milestone date at the milestone date.</td>
<td></td>
</tr>
<tr>
<td>26. CUSTOMER CONTACT TIME</td>
<td>Average hours per researcher spent in direct contact with external (or internal) customers.</td>
<td>PRD</td>
</tr>
<tr>
<td>27. PRESERVATION OF TECHNICAL OUTPUT</td>
<td>Percent of research projects outcomes captured in technical reports.</td>
<td>PRD</td>
</tr>
</tbody>
</table>
## Definitions for the Menu

<table>
<thead>
<tr>
<th>Metric</th>
<th>Definition</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>28. Efficiency of Internal Technical Processes</strong>&lt;br&gt;a. Project Assessment&lt;br&gt;b. Portfolio Assessment</td>
<td>The total cost of all commercially successful projects divided by the number of commercially successful projects.&lt;br&gt;The total R&amp;D budget divided by the number of projects with commercial output. Subdivide by projects of similar type (technical service, short-term, long-term) and use in conjunction with project value assessment.</td>
<td>PA: PRD</td>
</tr>
<tr>
<td><strong>29. Employee Morale</strong></td>
<td>Quantitative ratings of key aspects of employee satisfaction and morale as shown by direct employee survey.</td>
<td>PRD</td>
</tr>
<tr>
<td><strong>30. Goal Clarity</strong></td>
<td>Interval rating scale assessing the extent to which project performance objectives are clearly identified and understood by all participants on the project team.</td>
<td>PRD</td>
</tr>
<tr>
<td><strong>31. Project Ownership/Empowerment</strong></td>
<td>Interval rating scale assessing the extent to which participants feel they have the support and freedom they need to be successful in the project.</td>
<td>PRD</td>
</tr>
<tr>
<td><strong>32. Management Support</strong></td>
<td>Interval rating scale assessing the extent to which participants feel they have management's backing and an understanding that failure while learning will not be punished.</td>
<td>PRD</td>
</tr>
<tr>
<td><strong>33. Project Championship</strong></td>
<td>Percent of projects for which an effective project champion can be identified on the project team.</td>
<td>PRD</td>
</tr>
</tbody>
</table>